

Université de Montréal

Les Évaluations Formatives de Performances Cliniques (ÉFPC)

**Un exemple d'approche authentique pour le développement
des compétences dans l'enseignement supérieur**

par :

Diem-Quyen Nguyen

Département d'administration et fondements de l'éducation
Faculté des sciences de l'éducation

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)
en sciences de l'éducation
option mesure et évaluation

Juin 2006

© Diem-Quyen Nguyen, 2006



LB
5
U57
2007
V.027

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

**Les Évaluations Formatives de Performances Cliniques (ÉFPC)
Un exemple d'approche authentique pour le développement des
compétences dans l'enseignement supérieur**

Présentée par :

Diem-Quyen Nguyen

a été évaluée par un jury composé des personnes suivantes :

Michel D. Laurier
Président du jury
Université de Montréal

Bernard Charlin
Membre du jury
Université de Montréal

Jean-Guy Blais
Directeur de recherche
Université de Montréal

Marielle Simon
Examinatrice externe
Université d'Ottawa

Michel D. Laurier
Représentant(e) du doyen de la FES
Université de Montréal

Thèse acceptée le : _____



Résumé

Cette thèse doctorale, présentée sous forme de quatre articles, est une recherche de développement d'une méthode d'évaluation formative qui vise à développer des compétences cliniques à un niveau de formation post-graduée.

L'objectif général de la thèse était d'appliquer les fondements théoriques de l'évaluation formative des compétences dans la recherche d'une méthode qui vise spécifiquement le développement des compétences par l'évaluation. Chacun des articles fait partie du processus de développement et de validation d'une nouvelle méthode d'évaluation formative des compétences.

Le premier article est un essai qui compare les fondements théoriques entre l'approche par objectifs et celle par compétences à la lumière d'une révision des écrits. Nous avons mis en contraste les avantages et les limites de chacune des deux approches pour mieux faire ressortir les nuances. La formation clinique sert d'exemple de l'application pratique des deux approches.

Le deuxième article est une étude pour valider le contenu d'une nouvelle «échelle descriptive globale» qui sera intégrée à une méthode d'évaluation authentique nommée «Évaluation Formative des Performances Cliniques» (ÉFPC). Pour établir les standards de performance, le modèle «l'onglet» de Lewis a été adapté dans un contexte de création d'une nouvelle échelle d'évaluation. Douze professeurs cliniciens ont participé au processus d'établissement des standards. Un projet pilote a eu lieu pour tester le contenu auprès des utilisateurs.

Le troisième article rapporte les résultats d'une expérience de six mois de mise à l'essai pour vérifier si les ÉFPC permettent un développement véritable des performances cliniques. Treize résidents ont été évalués de façon aléatoire par six professeurs une fois par mois pendant ces six mois. Chaque session d'ÉFPC consiste à demander au résident

de résoudre un problème clinique d'un patient véritable après l'avoir interrogé, examiné et avoir répondu à ses questions. Par la suite, la rétroaction est ciblée sur les points de divergence entre l'auto-évaluation et l'évaluation des professeurs sur la performance du résident. L'analyse quantitative des résultats des performances au début et à la fin, combinée à l'analyse qualitative des commentaires écrits, ont été effectuées pour établir le profil de progression des résidents.

Finalement, le quatrième article étudie la composante «auto-évaluation» dans l'évaluation formative. Cette analyse a lieu grâce à l'étude de corrélation entre les scores de l'auto-évaluation et les scores de l'évaluation par les professeurs. De nouveau, l'analyse des commentaires écrits, combinée à l'analyse quantitative, permettra de comprendre comment le résident développe sa capacité d'auto-évaluation.

Les résultats de nos travaux tentent de montrer qu'une approche d'évaluation authentique respectant les principes fondamentaux de l'évaluation formative favorise le développement des compétences cliniques, même dans une courte période de temps.

Nous pouvons de plus conclure que pour développer l'auto-évaluation, une rétroaction ciblée par des professeurs compétents dans le domaine semble nécessaire. En effet, même si l'échelle descriptive globale est appréciée par les professeurs pour ancrer des points de discussion, elle ne permet une utilisation juste par les étudiants à s'auto-évaluer que si elle est combinée à une rétroaction.

Mots-clés

Recherche de développement – Enseignement professionnel – Évaluation formative – Approche authentique – Échelle descriptive globale – Auto-évaluation – Compétences –

Abstract

This doctoral thesis, presented in a series of four articles, represents a research on educational development of a formative assessment method that will help in the development of clinical competencies at the post-graduate level.

The general objective of the thesis is to apply the elemental theories of formative assessment in the search for a method whose aim is to help medical residents in developing their clinical competencies. Each of the four articles is a component in the developing and validating processes of the "Formative Clinical Performance Assessment" (Form-CPA) as an effective method to assess formatively clinical competencies.

The first article is an essay where the competency-based approach is compared to the objectives-based approach in education. The clinical training program provides an example of a practical application of both approaches in a program development.

The second article represents a study to validate a new global descriptive scale that will be incorporated into the formative authentic assessment method of clinical competencies, the Form-CPA. Twelve teachers participated in the standard setting procedure that is inspired from the "Lewis bookmark method ". The model was piloted with a small group of users (teachers and residents) testing the content aspect of this new scale.

The third article reports the results of the experimentation of the Form-CPA on a 6 teachers and 13 residents cohort over a six-month period. This step aims at finding out whether or not clinical competencies can be improved by the Form-CPA method. Once a month, each resident met the teacher who has been randomly assigned to him for a Form-CPA session. At this session, the resident was to perform a questionnaire and a physical examination on a real patient under direct observation. The resultant feedback was intended primarily to target the difference between the self-assessment and the teacher's

assessment of resident's performance. A combined quantitative and qualitative analysis was carried out to establish the progress profile of the residents.

The fourth article is a study about the self-assessment component of the Form-CPA. A correlational analysis, combined with a qualitative analysis, was performed to establish the baseline performance of the resident in self-assessment and his evolution during the six-month experimentation period.

Our results show that the Form-CPA is a valid and reliable method to assess formatively residents' clinical competencies. Significant improvements have been recorded during the six months of follow-up period. Our research seems to confirm the hypothesis that authentic assessment may be efficient in helping toward the development of complex professional competencies. The feedback component by competent teachers should be considered as an essential component if self-assessment skills are to be developed. Our findings seem to point out that the use of a global descriptive scale should be combined with an in-depth feedback from the teachers for an efficient development of self-assessment skill.

Keywords

Developmental research – Formative assessment – Professional training – Authentic assessment – Global descriptive scale – Self-assessment – Competence –

Table des matières

IDENTIFICATION DU JURY	II
RÉSUMÉ FRANÇAIS	III
MOTS-CLÉS	IV
ENGLISH ABSTRACT	V
KEYWORDS	VI
TABLE DES MATIÈRES	VII
LISTE DES TABLEAUX	XIV
LISTE DES FIGURES	XV
LISTE DES ANNEXES	XVI
DÉDICACE	XVII
REMERCIEMENTS	XVIII
AVANT-PROPOS	XIX
PRÉSENTATION DE LA THÈSE	1
INTRODUCTION	2
QUESTIONS DE RECHERCHES	5
MÉTHODOLOGIE	6
CHOIX DES PARTICIPANTS	8
DESCRIPTION DE LA MÉTHODE D'ÉVALUATION FORMATIVE DES PERFORMANCES CLINIQUES (ÉFPC)	9

PRÉSENTATION DES QUATRE ARTICLES	
Article 1.....	10
Article 2.....	10
Article 3.....	11
Article 4.....	11
ARTICLE 1	12
APPROCHE PAR OBJECTIFS OU APPROCHE PAR COMPÉTENCES : QUELLES IMPLICATIONS POUR L'ENSEIGNEMENT ET POUR L'ÉVALUATION DANS LA FORMATION CLINIQUE ?	
RÉSUMÉ	14
MOTS CLÉS	15
ABSTRACT	16
KEYWORDS	
I INTRODUCTION	17
II APPROCHE PAR OBJECTIFS	19
II.1 Définition et implications	19
II.2 L'évaluation dans l'approche par objectifs	22
II.3 Avantages et limites de l'approche par objectifs.....	23
III APPROCHE PAR COMPÉTENCES	25
III.1 Définition et implications	25
III.2 L'évaluation dans l'approche par compétences	33
III.3 Avantages et limites de l'approche par compétences.....	34
IV COMPARAISON DES DEUX APPROCHES PÉDAGOGIQUES	35
V ENSEIGNEMENT CLINIQUE : QUELLE APPROCHE?	39
VI CONCLUSION	43
RÉFÉRENCES	45

ARTICLE 2	52
MODÉLISATION D'UNE GRILLE D'ÉVALUATION FORMATIVE DES PERFORMANCES DANS UNE PERSPECTIVE D'APPROCHE PAR COMPÉTENCES : UN EXEMPLE DU PROGRAMME DE FORMATION POSTDOCTORALE EN MÉDECINE INTERNE	53
RÉSUMÉ	54
MOTS CLÉS	54
ABSTRACT	55
KEYWORDS	55
I. L'ÉVALUATION FORMATIVE DES COMPÉTENCES	56
I.1 Évaluation formative	56
I.2 Évaluation des compétences.....	60
I.3 Grille d'évaluation pour des performances	61
I.4 L'échelle de mesure.....	63
II. CONTEXTE ET CONTRAINTES DE L'ÉVALUATION FORMATIVE DES COMPÉTENCES CLINIQUES	66
III. DÉVELOPPEMENT DES ÉVALUATIONS FORMATIVES DES PERFORMANCES CLINIQUES (ÉFPC)	71

IV	CRÉATION DE LA GRILLE	74
IV.1	Validité et fidélité des mesures.....	74
	<i>IV.1.1 Validité et fidélité des mesures pour</i>	
	<i>des performances complexes</i>	76
IV.2	Étapes de développement d'une grille descriptive globale pour évaluer	
	des performances cliniques.	78
	<i>IV.2.1 La première étape : Décision sur le contenu de la grille</i>	78
	<i>IV.2.2 La deuxième étape : Établir la validité du contenu de la grille</i>	
	<i>auprès des experts</i>	79
	<i>IV.2.3 La troisième étape : valider la grille auprès d'un nombre restreint</i>	
	<i>d'utilisateurs</i>	85
IV.3	Résultats de la troisième étape.....	86
V.	DISCUSSION ET CONCLUSION	88
	RÉFÉRENCES	91

ARTICLE 3	102
L'ÉVALUATION FORMATIVE PAR UNE APPROCHE D'ÉVALUATION AUTHENTIQUE :	
RÉSULTATS DES ÉVALUATIONS FORMATIVES DES PERFORMANCES CLINIQUES DU PROGRAMME DE MÉDECINE INTERNE	105
RÉSUMÉ	106
MOTS-CLÉS	106
ABSTRACT	107
KEYWORDS	107
I. INTRODUCTION	108
II. MÉTHODOLOGIE	111
II.1 Description de la méthode « Évaluation Formative des Performances Cliniques » (ÉFPC)	111
II.2 Recrutement des professeurs, des résidents et des patients	113
II.3 Considérations éthiques.....	114
II.4 Plan de recherche.....	115
II.5 Plan d'analyse des données	115
III. ANALYSE DES DONNÉES ET RÉSULTATS	116
III.1 Profil des participants et des sessions d' ÉFPC.....	116
III.2 Consistance interne de la grille d' ÉFPC	118
III.3 Étude de la progression des performances	118

III.4	Analyse des annotations sur la grille d'ÉFPC	120
III.5	Perception des professeurs de l'ÉFPC	121
IV.	DISCUSSION ET CONCLUSION	125
	RÉFÉRENCES :	146
 <u>ARTICLE 4</u>		
DÉVELOPPEMENT DE LA CAPACITÉ D'AUTO-ÉVALUATION DES ÉTUDIANTS PAR UNE MÉTHODE D'ÉVALUATION AUTHENTIQUE : RÉSULTATS AVEC DES ÉFPC		151
	RÉSUMÉ	152
	MOTS-CLÉS	153
	ABSTRACT	154
	KEY WORDS	154
I.	INTRODUCTION	155
II.	MÉTHODOLOGIE	162
II.1	Description de la méthode d'ÉFPC	162
II.2	Le déroulement du protocole de recherche	163
III.	LE RECRUTEMENT DES RÉSIDENTS, DES PROFESSEURS ET DES PATIENTS	164

V. ANALYSE DES RÉSULTATS	165
V.1 Résultats de l'évolution des corrélations.....	165
V.2 Analyse qualitative des annotations	169
V.3. Analyse qualitative de la perception des résidents.....	170
 VI. DISCUSSION ET CONCLUSION	 173
 RÉFÉRENCES	 184
 CONCLUSION GÉNÉRALE	 188
 SOURCES DOCUMENTAIRES	 194

Liste des tableaux

ARTICLE I

Tableau 1 :	Comparaison des deux approche; par objectifs et par compétences	39
-------------	-----------------------------------------------------------------	----

ARTICLE 3

Tableau II :	Profil des professeurs-évaluateurs.	129
Tableau III :	Profil des résidents	130
Tableau IV :	Caractéristiques des sessions d'ÉFPC	131
Tableau V :	Comparaison des compétences au début et à la fin	132
Tableau VI :	Comparaison de l'évolution des sous-compétences	133
Tableau VII :	Analyse qualitative des annotations	134

ARTICLE 4

Tableau VIII :	Corrélation des performances au début et à la fin	177
TABLEAU IX :	Fréquence des performances selon les professeurs et les résidents	179
Tableau X :	Proportion de concordance des scores	180
Tableau XI :	Annotations et commentaires sur la grille par résidents et professeurs	182
Tableau XII :	Exemples des commentaires des résidents et des professeurs sur les forces et les faiblesses des résidents	183

Liste des figures

ARTICLE 3

Figure 1 :	Évolution du questionnaire du problème principal	135
Figure 2 :	Évolution du questionnaire sur l'impact psychosocial	136
Figure 3 :	Évolution de la technique de l'entrevue	137
Figure 4 :	Évolution de l'examen physique	138
Figure 5 :	Évolution de la discussion avec les patients	139
Figure 6 :	Évolution de la présentation de cas	140
Figure 7 :	Évolution du diagnostic différentiel	141
Figure 8 :	Évolution du plan d'investigation	142
Figure 9 :	Évolution du plan thérapeutique	143
Figure 10 :	Compétences cliniques	144
Figure 11 :	Compétence globale	145

ARTICLE 4

Figure 12 :	Évolution de la corrélation des performances	178
Figure 13 :	Proportion de la concordance des scores	181

Liste des annexes

Annexes

Annexe I

Évaluation formative des performances cliniques	II
Documents remis aux professeurs	II

Annexe II

Évaluation formative des performances cliniques	XII
Documents remis aux résidents	XII

Annexe III

Grille D'ÉFPC (version initiale)	XXIII
----------------------------------	-------

Annexe IV

Grille d'ÉFPC (3 niveaux)	XXIV
---------------------------	------

Annexe V

Grille d'ÉFPC (4 niveaux)	XXVIII
---------------------------	--------

Je dédie ce travail à :

- *Mes enfants, Claudia, Christina et Julian; la raison d'être même de ma vie*
- *Mon mari, Khue Nguyen, pour tous les bons et difficiles moments des dernières six années*
- *Mes parents décédés il y a six ans; sans eux, je ne serais jamais là où je suis actuellement*
- *Mes beaux-parents qui n'ont cessé de me supporter et encourager*
- *Aux treize étudiants et cinq professeurs, que je ne peux nommer, mais sans eux, ce travail serait beaucoup plus difficile*

Remerciements

Cette thèse ne serait jamais réalisée s'il n'y avait pas la présence et l'aide des personnes suivantes :

- Madame Huguette Bernard, qui a réussi à me ramener des grandes idées floues à un sujet clair et solide
- Monsieur Bernard Charlin, qui m'a aidé à clarifier mes idées par ses commentaires judicieux
- Monsieur Jean-Guy Blais, qui a accepté de prendre la relève et de me guider tout au long du processus de réflexion et de rédaction
- L'équipe du Service de médecine interne de l'hôpital Saint-Luc du CHUM, qui n'ont cessé de me soutenir et plus particulièrement les docteurs Patrick D'Amico, Jean-Victor Patenaude et Roch Parent, pour leurs encouragements constants.

Avant-propos

Au cours de mes quinze années d'implication à différents niveaux dans l'enseignement médical, ce qui me surprend toujours, c'est le dévouement des professeurs dans l'enseignement et la grande satisfaction que l'on retire au moindre remerciement de la part de nos étudiants.

Les milieux de l'enseignement et médical au Québec ont vécu des moments de bouleversements majeurs au cours des dernières dix années. Il y a eu, en même temps, d'extraordinaires explosions de nouvelles découvertes dans les théories cognitives de l'apprentissage, des pressions sociales importantes pour mieux préparer les étudiants à faire face au travail professionnel qui les attend. Plus cela progresse, plus on voit un rôle croissant des responsabilités professorales. Et pourtant, en même temps, les coupures dans les ressources à tous les niveaux obligent un réajustement et une réflexion sur l'efficience dans l'enseignement et sur le rôle de tout le système scolaire dans la formation des professionnels de demain.

En parallèle à ces bouillonnements, le milieu médical au Québec a aussi vécu d'extraordinaires bouleversements. L'explosion des connaissances, l'augmentation vertigineuse des besoins de services due au vieillissement de la population, l'exigence sociale d'une médecine basée sur les preuves demande beaucoup de temps à des professionnels dans le milieu médical. Les coupures budgétaires aux mêmes moments ont obligé tout le monde à «faire mieux avec ce qu'on a ...».

C'est dans ce contexte que je suis retournée à l'école. Plusieurs m'ont demandé pourquoi. Sur le coup, j'avais de la difficulté à répondre. Mais, il est devenu de plus en plus clair que pour faire mieux, il ne suffit pas d'y croire, mais il faut aussi **prouver** que c'est possible, que c'est faisable et que c'est pour le mieux.

Je cherche alors à mieux maîtriser l'art et la science de l'éducation, comme on m'a enseigné à maîtriser l'art et la science de la médecine. Je voulais aussi trouver un moyen, une méthode, une façon de donner le meilleur à nos étudiants. Et, lorsque j'observe l'extraordinaire savoir-faire des professeurs, je me suis dit : "*... il faut faire profiter de tout cela à nos étudiants, puisque demain, ce sera à leur tour de nous soigner et d'enseigner à nos enfants et petits-enfants ...*".

C'est alors le début, et aujourd'hui, l'aboutissement d'un extraordinaire voyage de découvertes et de réflexions à côté des étudiants et des professeurs des deux facultés de Sciences de l'éducation et de Médecine.

Introduction

L'enseignement, aux études supérieures, prend des dimensions particulièrement excitantes depuis les dernières cinquante années, et plus particulièrement ces quinze dernières années. Sous les pressions sociales des milieux de travail, les curriculums évoluent et mettent de plus en plus l'accent sur une meilleure préparation de l'étudiant à faire face aux demandes de son domaine de travail, à se maintenir au courant de nouvelles connaissances et de nouvelles façons de faire tout en s'intégrant harmonieusement à différents milieux de travail.

Nous parlons alors du développement des compétences professionnelles et, nous parlons aussi de l'imputabilité des programmes de formation envers la société et nos propres étudiants (Linn, 2000).

Le domaine de l'évaluation a toujours influencé profondément l'apprentissage et joué un rôle important dans l'enseignement. Popham (2005) a ainsi déclaré que : «... *l'évaluation efficace va améliorer l'enseignement ...*». Le concept évaluer pour mieux enseigner n'est pourtant pas un concept couramment accepté, Popham (2005) le reconnaît d'ailleurs : «...*les professeurs aiment enseigner mais aiment rarement évaluer ...*». Earl (2003) considère aussi que l'évaluation fait partie de l'apprentissage et que les professeurs ont la responsabilité d'évaluer pour mieux enseigner. Huba et Freed (2000) parlent d'un changement de paradigme et suggèrent une évaluation qui soit recentrée sur l'étudiant. Selon eux, ce changement est nécessaire si l'on veut améliorer la qualité de la formation des étudiants. Des éléments importants qui justifient ce changement, ils ont mentionné la motivation de l'étudiant pour qu'il devienne actif dans sa propre construction des connaissances. À cette fin, ils suggèrent de mettre l'accent sur la

transmission et l'utilisation efficace des connaissances en utilisant des problèmes réels ou émergents.

Le défi est alors de taille. L'enseignement professionnel adopte de plus en plus la formation sous forme de stage dans des milieux réels de travail pour préparer progressivement l'étudiant à son futur milieu de travail. Cette stratégie de formation a des avantages indéniables pour développer des compétences et une identité professionnelles mais pose aussi des défis pour enseigner et évaluer dans un contexte de formation imprévisible à plus d'un titre, tout en respectant les principes selon lesquels l'étudiant devrait être au centre des processus d'apprentissage. Dans ce contexte, le concept de « l'évaluation formative prend toute son importance ».

Selon Black et Williams (1998), l'évaluation formative est efficace à tous les niveaux et dans tous les contextes de formation puisque son rôle principal est d'améliorer l'apprentissage, en soulignant que la rétroaction sur la performance joue un rôle clé. La qualité de cette rétroaction semble être cruciale. Leur révision des écrits sur l'évaluation formative semble par contre démontrer une grande variabilité dans les approches pour évaluer et pour donner la rétroaction. De plus, les effets qui en découlent peuvent être positifs ou négatifs.

Des approches utilisées en évaluation, que ce soit à des fins sommatives ou formatives, notons l'approche émergente appelée évaluation authentique qui s'impose progressivement dans les curriculums adoptant une pédagogie basée sur le développement des compétences. Cette approche propose une évaluation dans un contexte proche de la vie réelle professionnelle et met l'accent sur l'intégration de tous les savoirs pour résoudre des problèmes complexes et mal définis. Les partisans de cette approche croient que l'appréciation, à la fois de la qualité du processus de résolution du

problème et du résultat même, favorise une meilleure connaissance des capacités de l'étudiant et favorise donc une meilleure rétroaction.

En formation clinique, très peu de recherches étudient spécifiquement les effets de l'évaluation formative sur le développement de compétences cliniques. Les études sur les méthodes d'évaluation sommative sont d'ailleurs très abondantes pour évaluer des performances cliniques. Il y a des études de validité, de fiabilité, de faisabilité et de standardisation pour de multiples méthodes d'évaluation, que ce soit de fin de cours, de stage, de session ou d'année. Peu d'études portent par contre sur des méthodes dont le but principal est d'assurer une progression du développement de compétences.

Le but de notre travail est donc de cerner les aspects de l'évaluation des compétences pour mieux comprendre les enjeux importants dont il faut tenir compte lors de l'élaboration et de la validation d'une méthode d'évaluation formative des compétences cliniques.

Questions de recherches

La question générale de recherche se pose ainsi :

Est-ce que l'évaluation authentique dans une perspective d'évaluation formative, au niveau des études supérieures à vocation professionnelle, notamment en médecine interne, favorise efficacement le développement des compétences cliniques ?

Nous divisons cette question globale en quatre questions spécifiques :

1. Quelles sont les différences entre l'approche basée sur les objectifs et celle basée sur les compétences ?
2. Puisque l'élaboration d'une méthode d'évaluation comprend la création d'un outil de recueil de données, comment doit-on construire une grille d'évaluation formative des performances dans une perspective d'approche par compétences ?
3. L'évaluation formative des performances cliniques (ÉFPC) peut-elle contribuer à évaluer la progression des résidents, dans un contexte de formation entièrement composée de stages où aucun des résidents n'a le même cheminement ?
4. Si l'auto-évaluation est une composante essentielle aux compétences professionnelles, est-ce que la méthode d'ÉFPC en favorise le développement ?

Méthodologie

Notre recherche qui a pour but de développer une méthode d'évaluation formative des compétences cliniques a duré 18 mois. Elle est divisée en trois principales phases : l'analyse de besoins éducatifs dans un curriculum basé sur l'approche par compétences, la conception de la méthode d'ÉFPC avec la modélisation de la grille d'évaluation et, la mise à l'essai avec l'évaluation des résultats. Cette démarche s'inspire de celle proposée par Van der Maren (1999) sur la mise au point d'activité et de matériel pédagogiques.

Au cours de la première phase, nous avons fait une révision approfondie des publications qui portent sur les approches pédagogiques basées sur les objectifs et les compétences pour cerner les caractéristiques des deux approches, leurs implications pédagogiques, leurs avantages et leurs limites. Cette révision a été effectuée dans la foulée des révisions des curriculums où l'adoption de l'approche par compétences est devenue la norme dans les programmes nord-américains de formation médicale.

De cette analyse, il nous est devenu apparent que le champ de l'évaluation formative des compétences cliniques requiert une plus grande attention dans la recherche pour s'assurer d'un développement progressif des compétences. La deuxième phase, la conception de la méthode des ÉFPC, a eu lieu de février à juin 2004. Il s'agit de la conception des ÉFPC et de la création de la grille d'évaluation qui servira comme principal outil de mesure. Au cours de cette phase, nous avons validé le contenu de cette grille et établi des standards de performance avec la participation de douze professeurs-cliniciens. Ces deux processus s'inspirent de la méthode « signet » pour l'établissement des standards de Lewis (1998).

Au cours de cette phase, nous avons élaboré une grille d'évaluation de type « échelle descriptive globale » intégrée à une méthode d'évaluation authentique. Cette

méthode nommée ÉFPC est composée d'entrevues cliniques avec de véritables patients qui sont suivies d'une rétroaction basée sur une combinaison de l'évaluation par les professeurs et de l'auto-évaluation. Une première mise à l'essai auprès des utilisateurs (cinq professeurs cliniciens et douze résidents) a eu lieu en janvier 2005 et a permis de raffiner la grille et de peaufiner les stratégies d'implantation.

La troisième phase est la mise à l'essai avec une cohorte de six professeurs cliniciens et de treize résidents du programme de médecine interne dans une étude longitudinale de six mois pour vérifier le développement des compétences cliniques en utilisant les ÉFPC.

Choix des participants

Douze des vingt professeurs sollicités ont participé à la phase de modélisation de la grille d'évaluation qui est utilisée lors des ÉFPC. Ils ont accepté, après avoir été invités sur la recommandation du comité de programme de médecine interne. Les critères de sélection ont été le domaine de spécialisation, la familiarité avec le programme de formation des études spécialisées en Médecine Interne, et leur réputation de dévouement à l'enseignement. Huit ont décliné l'offre, en invoquant un manque de temps. Les douze qui ont accepté proviennent de diverses spécialités médicales: médecine interne, neurologie, pneumologie, endocrinologie, cardiologie, soins intensifs et gastro-entérologie. Ils avaient tous un minimum de cinq années d'expérience dans l'enseignement clinique.

Au cours de la phase de la mise à l'essai, cinq professeurs cliniciens ont accepté de participer. Nous les avons choisis en nous basant sur leur expérience antérieure avec l'observation directe, leur réputation de dévouement (afin d'éviter des abandons au cours de l'expérimentation), leur connaissance du programme de médecine interne et leur domaine d'expertise clinique. Quant aux résidents, treize ont accepté de participer après deux sollicitations par lettre circulaire auprès de tous les résidents des trois premières années du programme de Médecine Interne de l'Université de Montréal.

Les patients ont été recrutés dans le bassin des patients des professeurs, en leur indiquant clairement le but académique de l'activité.

Description de la méthode d'Évaluation Formative des Performances Cliniques (ÉFPC)

Les ÉFPC visent à évaluer de façon structurée et formelle les compétences cliniques des résidents pour les aider à progresser.

Chaque session d'ÉFPC est composée de deux parties, la première étant la performance clinique du résident et la deuxième étant consacrée à la rétroaction. Elle devrait durer 60 minutes. Au cours des premières trente minutes, le professeur observe le résident interroger un patient sur son problème médical principal, l'examiner de façon pertinente à ce problème et répondre aux questions du patient.

Par la suite, en l'absence du patient, le résident résume les données cliniques et présente un diagnostic différentiel, un plan d'investigation et, le cas échéant, un plan de traitement. Pour s'assurer du caractère authentique de l'évaluation, le patient est un véritable patient avec un problème médical véritable, susceptible de consulter le résident dans sa vie professionnelle ultérieure.

La grille d'évaluation qui sert à recueillir les données est de type d'échelle descriptive globale. Le contenu de cette grille ainsi que la description de chaque niveau de standard de performance sont préalablement validés.

Les professeurs participant à l'étude reçoivent un document qui inclut la description de la méthode des ÉFPC, du processus de rétroaction et une copie de la grille d'évaluation. Ce document se trouve à l'annexe I. Les résidents ne reçoivent quant à eux que la description des ÉFPC et la grille d'évaluation (alinéa, annexe II).

Présentation des quatre articles

Cette recherche est présentée sous forme de quatre articles, chacun décrivant une étape cruciale du développement de la méthode des ÉFPC comme méthode d'évaluation formative des compétences professionnelles complexes telles que les compétences cliniques.

Le premier article a été écrit dans l'optique d'être soumis pour publication à la revue « Pédagogie médicale ». Le deuxième est un article à visée plus technique et sera soumis à une revue spécialisée telle que « Mesure et Évaluation en Éducation ». Les deux derniers articles seront éventuellement traduits en anglais pour être proposés à la revue « Academic Medicine » et « Advances in Health Education ».

Article 1

L'article numéro 1, intitulé « Les implications de l'approche par objectifs ou de l'approche par compétences dans l'enseignement et l'évaluation dans la formation clinique », vise à réviser les publications portant sur les deux approches, et d'en dégager les grandes caractéristiques pour mieux comprendre les avantages et les limites de chacune de ces approches.

Cet article est divisé en trois sections: la première tente de comprendre les fondements théoriques de l'approche par objectifs, la deuxième tente de décortiquer le concept de compétence et ses implications dans l'enseignement et l'évaluation, la troisième est surtout un exercice de compréhension et d'application des deux approches dans la formation clinique.

Article 2

Le deuxième article, « Modélisation d'une grille d'évaluation formative des performances dans une perspective d'approche par compétences: un exemple du

programme de formation post-doctorale en Médecine Interne » est une étude de validation du contenu d'une grille d'évaluation qui servira à recueillir des données sur la performance des résidents lors des entrevues médicales. Le processus suivi est un processus argumentatif des différents aspects théoriques qui jouent un rôle important dans la création d'une grille, allant de la définition de l'évaluation jusqu'aux processus d'établissement de standards de performances, en passant par l'analyse des modèles de grille d'évaluation.

Article 3

Tel que son nom l'indique, l'article 3, « L'évaluation formative par une approche authentique: résultats des ÉFPC du programme de Médecine Interne », rapporte les résultats de l'expérimentation des ÉFPC après six mois avec six professeurs et treize résidents. Cet article tente de vérifier l'hypothèse selon laquelle une méthode d'évaluation authentique qui tient compte de tous les principes de l'évaluation formative, permet véritablement de développer des compétences cliniques, telles qu'elles sont jugées par les professeurs.

Article 4

Finalement, l'article 4 traite du rôle de l'auto-évaluation dans le développement de compétence. Cet article cherche surtout à prouver l'hypothèse selon laquelle, l'évaluation authentique avec l'utilisation d'une grille globale descriptive favorise le développement de l'auto-évaluation, toujours dans une perspective d'évaluation formative. Dans cet article intitulé « Développement de l'auto-évaluation des étudiants par une méthode d'évaluation authentique: résultats avec des ÉFPC », nous mettons l'accent sur la comparaison des perceptions qu'ont les résidents et les professeurs sur les performances cliniques de ces résidents.

Article 1

**Approche par objectifs ou approche par compétences :
quelles implications pour l'enseignement
et pour l'évaluation dans la formation clinique ?**

Diem-Quyen Nguyen

Département d'administration et fondements de l'éducation

Université de Montréal

© 2006

Résumé

Les courants dominants en pédagogie médicale ont évolué constamment ces cent dernières années. Ainsi, à partir des curriculums axés sur le contenu, on a vu au cours des années 1970 l'utilisation élargie de l'approche par objectifs dans les programmes de formation. Mais, depuis 1990, l'approche par compétences s'impose progressivement dans plusieurs programmes médicaux du niveau premier cycle à ceux des cycles supérieurs.

Ce présent essai a pour but d'analyser, par une revue des publications pédagogiques, les caractéristiques des deux approches, ainsi que leurs implications dans l'enseignement et dans l'évaluation des apprentissages.

Notre analyse suggère que les différences fondamentales entre l'approche par objectifs et l'approche par compétences dans les curriculums sont de trois ordres. La première approche est centrée sur le professeur dans toutes les démarches de planification, de prestation et de l'évaluation des apprentissages, tandis que dans la deuxième approche l'étudiant est responsable du développement de ses propres compétences sous le guide des professeurs. L'approche par objectifs est par ailleurs caractérisée par la rigueur avec laquelle les objectifs sont clarifiés et classés selon des taxonomies préétablies et acceptées, tandis que l'approche par compétences met l'accent sur l'intégration des connaissances et des habiletés psychomotrices et affectives pour résoudre des problèmes dans des contextes variés. Finalement, l'évaluation selon l'approche par objectifs est basée sur l'évaluation des résultats observables, tandis que l'évaluation des compétences cherche à évaluer le processus qui a donné le résultat observable ainsi que l'appréciation de la qualité du résultat même.

Les avantages et les limites de ces deux approches sont discutés et des perspectives de recherche sont proposées.

Mots clés

Objectifs – Compétences – Éducation médicale – Curriculum – Éducation post-graduée – Éducation prégraduée – Évaluation.

Abstract

Medical education has been evolving over the last hundred years, from content-based curriculum through the instructional objectives-based approach since 1970. In the last 15 years, however, competency-based approach is taking over postgraduate medical education as well as the undergraduate medical curriculum.

The purpose of this essay is to analyze, through a review of educational literature, the characteristics of both approaches, as well as the educational impacts of these two approaches on teaching and assessment.

The results of this analysis shed light on two educational approaches. The objectives-based approach emphasizes the need to have clear, and specific instructional objectives, is centered on the teacher's role in instructional design, and assesses only the observable results and behaviours. The competency-based curriculum, in contrast, places significant importance on the integration of all the cognitive and non-cognitive skills to solve each problem in its context. This approach, furthermore, is centered on the student in his development of necessary competencies. The assessment, according to the competency-based approach, seeks to draw attention to the visible results as well as the process that brings about these results.

Advantages and limits of both approaches are discussed and research-perspectives are proposed.

Keywords

Medical Education – Postgraduate – Undergraduate – Curriculum – Instructional objectives – Competency-based assessment.

I Introduction

Depuis les dernières quinze années, les articles de revues et les livres en pédagogie médicale utilisent abondamment le mot «compétences» dans la formation, mais dans les livres publiés sur la planification de l'enseignement, le mot «objectifs» y est largement utilisé. Y a-t-il une différence entre ces deux approches pédagogiques sous-jacentes à ces mots ? Quelles sont les implications d'une approche par objectifs par rapport à une approche par compétences ? Pourquoi l'approche par objectifs, si populaire au cours des trente dernières années, est-elle critiquée et remise en question ? Pourquoi ce besoin d'une nouvelle approche ?

L'éducation médicale a vécu d'importants changements au début du 20^e siècle. D'une formation plus ou moins aléatoire basée principalement sur le compagnonnage, l'on est passé à la création des systèmes d'écoles de médecine affiliées à des hôpitaux. La formation très variable qui dépendait d'un seul maître est devenue plus standardisée aux États-Unis suite aux recommandations de Flexner (1910). L'accent était progressivement mis sur le contenu scientifique qui doit être acquis dans des laboratoires et dans les milieux hospitaliers.

Nanson (1975), professeur en chirurgie à l'Université Auckland, notait dans un article de réflexion sur 100 ans de l'évolution de l'éducation médicale à cette université, qu'un étudiant au début des années 70 devait maîtriser jusqu'à environ 32 matières scientifiques. Selon lui, la «quantité de matière en médecine paraissait épouvantable pour un étudiant néophyte et semblait être insurmontable pour d'autres étudiants, même les plus intrépides ».

Au cours des années 1960, le contact avec des milieux de l'éducation et de la sociologie a exercé une grande influence sur le milieu de l'éducation médicale. Les

pédagogues fréquemment cités sont Benjamin Bloom, Nathaniel Cantor, Stephen Abrahamson et Georges Miller. La formation basée sur le contenu commence à être remise en question et des curriculums basés sur l'application plus rigoureuse des principes pédagogiques, allant de la planification jusqu'à l'évaluation du programme de l'enseignement, font leur apparition. Plusieurs programmes de formation médicale se sont adaptés à l'approche par objectifs lors de la révision de leur curriculum (Ferland, 1987). Pour tous, l'attrait principal de cette approche était l'obligation de clarifier les intentions et de spécifier les performances désirées (Pugh, 1975). Mais, après un engouement considérable pour l'approche par objectifs dans la formation médicale, au cours des années 1970 et 1980, les éducateurs ont été confrontés aux difficultés de son application sur le terrain. Cette approche exige en effet une élaboration des objectifs spécifiques parfois trop détaillés. La prise de conscience des limites associées à cette approche soulève des questions sur la fragmentation des domaines de compétences et du degré de spécificité détaillée des objectifs de l'enseignement (Huba & Freed, 2000; Voorhees, 2001).

L'explosion des connaissances médicales, les pressions sociales pour mieux former les médecins ainsi que les développements en psychologie cognitive et son application en éducation introduisent alors le concept de compétences dans la formation médicale. On veut des médecins compétents dans tous les domaines et non seulement des médecins savants. Les compétences de type « apprendre à apprendre », « résolution des problèmes »... commencent à apparaître dans les programmes de formation médicale. Des méthodes telles que l'apprentissage par problème, l'introduction précoce à l'enseignement clinique par la formation des stages, sont progressivement adoptées par un nombre croissant d'universités (Des Marchais, 1996; Norman, 2002).

Cet article tente de clarifier, dans un premier temps, le concept d'objectifs pédagogiques et revoir les caractéristiques de l'approche par objectifs, son implication dans l'enseignement et dans l'évaluation, ainsi que ses avantages et ses limites.

Dans un deuxième temps, pour faire un parallèle avec l'approche par objectifs, le concept de «compétences » est défini. Ses caractéristiques et les implications sur le plan de l'enseignement et de l'évaluation sont discutées. Par la suite, une comparaison des deux approches pour faire ressortir les avantages et les limites respectives permettra de comprendre l'utilité de l'une ou de l'autre approche dans le développement d'un curriculum, d'un programme ou tout simplement d'un cours.

Finalement, un exemple de l'application de ces deux approches dans la formation clinique sera présenté.

II Approche par objectifs

II.1 Définition et implications

Selon de Landsheere (1984), l'engouement pour l'utilisation des objectifs pédagogiques provient d'un besoin profondément ressenti dans tous les milieux d'éducation à la fin des années 1950 et surtout au cours des années 1960. Confrontés à l'augmentation exponentielle des connaissances et du nombre d'étudiants, les éducateurs ont réalisé qu'une planification de l'enseignement selon une démarche systémique, comprenant une analyse des besoins, la détermination des objectifs d'apprentissage, le choix des méthodes d'enseignement et l'évaluation de l'atteinte des objectifs, serait plus efficace (Brien, 1981). Cette planification de l'enseignement repose sur ce que l'on appelle « des objectifs pédagogiques ».

Selon Bloom (1971), l'objectif pédagogique est une déclaration claire de ce que l'action éducative doit amener comme changement chez l'étudiant. Mager (1975) précise que les objectifs décrivent un ensemble de comportements que l'étudiant doit maîtriser. L'objectif devrait donc décrire ce qui est désirable et non le processus d'enseignement. Gagné (2005) ajoute que les objectifs devraient être considérés comme une déclaration

explicite des résultats désirés au terme d'un processus d'enseignement. Les objectifs d'apprentissage précisent donc des changements durables et désirables chez l'étudiant et qui surviennent pendant ou suite à une situation pédagogique (Legendre, 1993).

Les techniques de rédaction des objectifs ont été, par la suite, l'objet de plusieurs propositions et débats. Mais, de Landsheere (1984) nous a donné l'une des classifications des plus fonctionnelle. L'auteur reconnaît que tous les objectifs n'ont pas le même degré de précision et propose de les classer à trois niveaux : les buts, les objectifs généraux et les objectifs spécifiques. Ainsi, la rédaction des objectifs va de déclarations très générales, constituant des buts de l'éducation, aux objectifs généraux appliqués à de grandes sections de programme ou des cours. Ces objectifs généraux classifient les différents types de comportements selon des taxonomies d'apprentissage, dont celle de Bloom (1956) est la plus connue. Elle divise les capacités humaines en trois domaines : le cognitif, le psychomoteur et l'affectif. Finalement, les objectifs spécifiques, selon Mager (1975), Gagné (1974) et de Landsheere (1984), servent à spécifier des conditions de réalisation de l'apprentissage. Pour cela, la tâche devant être maîtrisée, le comportement observable issu de la formation, la situation et les conditions au cours desquelles ce comportement devrait être observé, ainsi que les critères qui serviront à l'évaluation, devront être spécifiés.

Toute cette démarche, allant de déclarations très générales reflétant souvent les besoins sociétaux dictés par des grandes institutions et organismes académiques (Harden, 1999 et Gagné, 2005), vers des objectifs spécifiques extrêmement détaillés, doit faciliter la planification des activités d'apprentissage et d'évaluation des professeurs. L'approche par objectifs répond alors aux besoins de ces professeurs qui désirent accorder leurs activités d'enseignement aux besoins de la société. Selon les auteurs tels que Bloom (1971), Gagné (1974), Mager (1975) et Scallon (1988), l'apprentissage est défini et évalué par des comportements observables puisque le résultat observé découle

directement de l'enseignement. Nous pouvons alors constater une importante influence du paradigme behavioriste selon lequel seuls les résultats observables devraient être évalués.

Gagné introduit en 1985 la notion des conditions d'apprentissage. Ceci constitue le début de l'application des connaissances développées en psychologie cognitive au domaine des sciences de l'éducation. L'apprentissage dépasse selon lui les simples principes de stimuli et de renforcement favorisés par la psychologie behavioriste. Gagné soutient que les différents stimuli et les rétroactions qui en résultent constituent l'ensemble des événements d'apprentissage qui favorisent l'intégration de l'information dans la mémoire et que c'est cet ensemble d'événements qui constitue les conditions d'apprentissage. Toujours selon Gagné (1985), les conditions de l'apprentissage devraient être précisées (conditions internes avec des préalables nécessaires incluant la motivation et conditions externes mises en place par le professeur) pour aider l'étudiant à maîtriser l'objectif.

L'exemple suivant permet de comprendre le concept des conditions internes et externes d'apprentissage. Pour apprendre à résoudre des problèmes de paralysie, l'étudiant devrait déjà être capable d'utiliser les règles et les informations de base sur l'anatomie du système nerveux et sur la physiologie de ce même système, ce qui constitue des conditions internes. Les conditions externes sont assurées par le professeur qui propose des problèmes diversifiés de paralysie. La résolution de ces problèmes fera appel aux règles et aux connaissances. Ces problèmes peuvent être des problèmes écrits sur papier ou des problèmes véritables et ce sont des conditions externes d'apprentissage (Gagné, 2005; adaptation libre de la page 73).

II.2 L'évaluation dans l'approche par objectifs

L'évaluation dans l'approche par objectifs met un accent important sur le résultat observable sans vraiment chercher à savoir comment ce résultat a été obtenu. Les professeurs devraient énoncer les critères de maîtrise et les conditions de réalisation des objectifs spécifiques quand ils élaborent des outils d'évaluation. Comme l'atteinte des objectifs de niveau taxonomique plus simples est plus facile à évaluer que le degré d'atteinte des objectifs de niveau taxonomique plus élevé, un grand accent a été mis sur l'évaluation des connaissances des faits, des règles, des lois et des principes, ce qui explique l'utilisation étendue en médecine des examens écrits et des examens ayant des items à choix multiples pour réponses.

Ces types d'examens jouissent d'un certain nombre d'avantages dont notamment celui de pouvoir être administré à de grands nombres d'étudiants, d'être corrigé de façon automatisée et d'avoir un niveau de fiabilité élevé (Downing, 2002; Schuwirth, 2004; Garibaldi, 2002). Ils sont par contre sujets à plusieurs critiques. La plus constante est que les activités évaluatives d'une approche par objectifs finissent souvent par évaluer des objectifs mineurs parce qu'ils sont faciles à évaluer, négligeant parfois les objectifs plus importants (Huba & Freed, 2000; Popham, 2005). Ainsi, pour des objectifs cognitifs de niveau taxonomique plus élevé, tels que l'analyse, la synthèse et l'évaluation, les difficultés de l'évaluation selon l'approche par objectifs deviennent plus apparentes.

Puisqu'il est parfois difficile de spécifier des critères de maîtrise et des conditions de réalisation, l'atteinte des objectifs affectifs et psychomoteurs n'est de plus pas évaluée aussi rigoureusement que celle des objectifs cognitifs. Ces évaluations peuvent donner un message négatif à l'étudiant qui ne perçoit pas ces domaines comme importants puisqu'ils ne sont pas évalués avec autant de rigueur (Williams, 1982; de Landsheere, 1984; Cushing, 2002).

L'exemple suivant permet de mieux comprendre les difficultés potentielles de l'évaluation des comportements complexes selon l'approche par objectifs. Faire une histoire de cas est l'un des objectifs principaux de la formation clinique. Dans le cas d'un patient présentant une douleur thoracique, l'étudiant doit interroger le patient sur les détails de base des caractéristiques de la douleur (quand, comment, ...) mais aussi sur les circonstances de la douleur (des facteurs aggravants, soulageants ...). Si l'on doit évaluer l'étudiant à ce sujet, il faut alors préciser les conditions de réalisation de cette histoire de cas comme préciser si le patient a été vu en cabinet privé, à la salle d'urgence ou lors de l'hospitalisation dans une unité de soins spécialisés. Si nous voulons de plus vérifier la capacité de synthèse et d'analyse de l'étudiant, il faut même ajouter encore plus de détails. Cet exemple illustre la difficulté de tout préciser et tout spécifier dans des objectifs spécifiques tel que le veut l'approche par objectifs en formation clinique.

II.3 Avantages et limites de l'approche par objectifs

L'approche par objectifs a plusieurs avantages dont l'un des principaux est qu'il faut expliciter par une déclaration détaillée les objectifs de l'enseignement. Cela facilite par la suite l'établissement de la congruence entre les objectifs recherchés et les activités éducatives proposées (de Landsheere, 1984). Il est important de souligner que la tâche des professeurs en est facilitée. Précisons que cette approche existe depuis presque un demi-siècle, et que plusieurs méthodes d'évaluation élaborées selon cette approche telles que les ECOS et les tests avec des questions à choix multiples ont été raffinées et que leur niveau de fiabilité et de validité est élevé, surtout en ce qui concerne les examens sommatifs à enjeu important (Downing, 2002; Reznick, 1996).

Par contre, lors de l'application pratique de cette approche, l'influence considérable du behaviorisme qui favorise les résultats observables est indéniable et que dans certaines situations cela est considéré comme une source de problèmes. Ainsi, les

professeurs sont souvent confrontés au dilemme du degré de spécificité des objectifs au niveau opérationnel. En 1977, dans un éditorial, le "General Medical Council" donnait comme exemple l'impressionnant ouvrage de 500 pages sur les objectifs d'apprentissage de l'école de médecine Abraham Lincoln et s'interroge déjà sur la pertinence de cette démarche (General Medical Council, 1977). Dans le même ordre d'idées, Mc Avoy, dans un article de 1985 où il encourageait l'utilisation des objectifs dans l'enseignement en médecine, mettait en garde les utilisateurs contre le danger de fragmentation lors de la spécification des objectifs pédagogiques. Sa crainte portait surtout sur le caractère inflexible dans l'énonciation des objectifs et sur le danger de se concentrer sur les « trivialités » et de perdre ainsi l'occasion d'ouvrir l'esprit vers ce qui est important à enseigner et à apprendre.

Harden (2002) a résumé une autre limite de l'approche par objectifs : il s'agit de la division artificielle des domaines d'apprentissage: le cognitif, le psychomoteur ou l'affectif. Cette division originellement, considérée comme ayant l'avantage important d'être structurée et rigoureuse dans l'énonciation des résultats de l'enseignement, reflète mal toute la nature complexe de l'apprentissage dans les domaines professionnels dont la médecine. Selon lui, à un niveau avancé de formation, l'étudiant doit être plus que simplement « bien informé » et l'intégration des connaissances, des attitudes et des habiletés psychomotrices pour agir comme médecin devient importante. Cette division en trois domaines favoriserait difficilement l'intégration désirée.

Une autre critique de l'approche par objectifs concerne surtout l'évaluation formative. L'approche par objectifs dans une perspective behavioriste propose d'effectuer une évaluation formative à partir des profils d'erreurs et de réussites observées lors des activités d'évaluation (Scallon, 1988). Mais, cette approche cause certains malaises dans le milieu de l'éducation. Par exemple, Tardif (1993) avance que si nous pouvons comprendre le processus cognitif de l'étudiant, il sera plus facile de l'aider à réajuster le

tir plutôt que d'attendre l'établissement des profils à partir des réussites et des erreurs commises et observées lors des activités d'évaluation. Cette vision est partagée de plus en plus par des auteurs ayant étudié la question de l'évaluation formative des apprentissages comme Huba et Freed (2000) et Scallon (2000). On retrouve le même écho dans les écrits en éducation médicale (Ben-David, 1999; Holmboe, 1996). Pour tous, l'évaluation du processus qui aboutit à un résultat est toute aussi importante que le résultat lui-même puisqu'elle permet d'aider l'étudiant à progresser.

L'émergence du concept de compétence et l'adoption progressive de l'approche par compétences dans plusieurs programmes de formation universitaire est alors vue comme une solution aux problèmes éprouvés avec l'approche par objectifs.

III Approche par compétences

III.1 Définition et implications

Le courant de formation basée sur les compétences a commencé à être diffusé en formation professionnelle universitaire au cours des années 1980 aux États-Unis (Tillema, 2000; Ewell, 1985) et il a pris de l'ampleur ces dernières années. Pour mieux comprendre ses implications dans l'enseignement et l'évaluation de l'approche par compétences, voyons d'abord ce qu'est une compétence et toutes ses composantes qui font partie de sa définition.

Voorhees (2001), se penchant sur les modèles d'apprentissage basés sur les compétences dans les études supérieures aux États-Unis, a proposé une définition de la compétence, en citant le "US Department of Education". Selon lui, la compétence est une intégration des habiletés, des connaissances et des capacités nécessaires à l'accomplissement d'une tâche spécifique. Il estime que l'un des éléments importants à la

base de toute activité d'apprentissage dans le développement des compétences est constitué par des traits et des caractéristiques propres à chaque étudiant (provenant de son bagage génétique, de son milieu familial et social) et qui influenceront l'acquisition des habiletés, capacités et connaissances en diverses situations d'apprentissage. L'intégration des connaissances, des habiletés psychomotrices et affectives au cours des processus d'apprentissage et la construction des savoirs lors des expériences d'apprentissage permettraient le développement des compétences. Ces compétences peuvent être évaluées, par des démonstrations de performances. Voorhees souligne ainsi l'aspect individuel sous-jacent à l'apprentissage et la notion de l'intégration des savoirs nécessaires au développement des compétences.

La définition de Voorhees, même si elle a l'avantage d'insister sur l'importance de l'intégration des capacités, des connaissances et des habiletés dans le développement des compétences, ne permet pas de comprendre ce qui distingue un médecin compétent en chirurgie d'un autre compétent en médecine générale, ni la raison pour laquelle un médecin endocrinologue est considéré comme expert en diabète et n'est que compétent en ostéoporose.

Des publications en pédagogie dans le milieu francophone apportent certaines clarifications importantes. Ainsi, Roegiers a cherché à savoir en 2001 comment développer les compétences dans l'enseignement. Il a alors tenté de faire la différence entre les objectifs, les capacités et les compétences. Selon lui, les objectifs expriment une intention d'amener un étudiant à exercer ses capacités sur un contenu spécifique, tandis que la capacité est tout simplement l'aptitude à faire quelque chose. Les compétences sont quant à elles liées à la tâche à accomplir et font appel à un exercice de jugement dans le choix de l'application des connaissances nécessaires pour accomplir efficacement une tâche, en tenant compte du problème en question (simple ou complexe) et du contexte (fictif ou réel) dans lequel l'action se déroule. Roegiers (2001) introduit ainsi les notions

de contexte et de jugement dans la définition de compétence et l'on retrouve de nouveau l'insistance sur la notion d'intégration des savoirs multidisciplinaires et multidimensionnels.

L'exemple suivant permet de mieux saisir les nuances de ces notions. Un étudiant de première année de médecine devrait être capable de faire un recueil de données cliniques et l'un de ses objectifs spécifiques sera de réaliser un recueil de données avec un patient âgé présentant un problème de douleur thoracique. Pour être déclaré comme compétent dans le domaine du recueil de données cliniques, le même étudiant devrait pouvoir faire un recueil de données cliniques avec des patients jeunes ou âgés, faciles ou difficiles à interroger sur des problèmes cliniques divers, quel que soit le contexte clinique (cabinet privé, urgence ...).

Perrenoud (1997) souligne quant à lui qu'une compétence n'est développée et stabilisée que lorsque l'individu peut mobiliser ses connaissances selon des schémas constitués en dépassant le stade du tâtonnement réflexif pour résoudre des problèmes dans différentes situations de la même famille de problèmes. En suivant cette logique, un médecin qui peut résoudre un problème de douleur thoracique, quelle que soit la provenance de la douleur, les caractéristiques du patient et du contexte clinique, est compétent pour résoudre des problèmes de douleur thoracique. Il a donc un jugement transférable d'une situation à d'autres situations de douleur thoracique.

Le Boterf (1995 & 2000) introduit la notion de niveau de développement des compétences en passant d'un état de novice à un niveau de professionnel compétent puis à celui de l'expertise professionnelle. Selon lui, l'individu engagé dans une profession se construit des savoirs intégrés et toutes ces étapes que nous venons d'énumérer sont essentielles dans le développement des compétences. Le parallèle de ce concept nous rappelle alors la base de la construction des connaissances selon la théorie piagétienne où l'on assiste au développement de la pensée abstraite chez les adolescents et les adultes,

après les stades sensori-moteurs des bébés et la pensée concrète des enfants. En formation médicale, Leach (2001) dans un article de réflexion sur la façon d'améliorer les soins offerts aux patients a proposé le même concept selon lequel des soins de qualité offerts aux patients doivent passer par le développement des compétences des professionnels qui les prodiguent. Ainsi, en médecine, l'étudiant représente le niveau de « novice ». Lorsqu'il a fini sa formation et commence sa pratique professionnelle, il atteint le niveau de « professionnel compétent ». Lorsqu'un médecin développe un champ d'excellence reconnu par les pairs et qu'on fait appel à lui en situations problématiques, il est alors reconnu comme « l'expert » dans ce domaine.

En résumé, selon Louis & Hensler (2003), on peut dégager trois grandes composantes suivantes de la compétence : (1) elle implique des actions efficaces en référence à des indicateurs de qualité pré-établis; (2) elle résulte de l'intégration des savoirs de plusieurs domaines (cognitif, affectif et psychomoteur) et; (3) elle se démontre dans des situations de problèmes complexes et nouveaux.

Ces éléments permettent de proposer une définition plus spécifique pour comprendre le concept de « compétence » et de proposer des pistes de réflexion sur la façon d'aider les étudiants à développer leurs compétences. Selon nous, **la compétence est une disposition qui permet la manifestation des capacités stabilisées d'un individu à mobiliser ses connaissances cognitives, ses habiletés affectives et psychomotrices de façon intégrée, en faisant des choix pour résoudre efficacement des problèmes complexes dans des familles d'opérations.**

Quelles implications découlent de cette définition dans l'enseignement et l'évaluation ?

Selon Scallon (2004), c'est dans l'action que les compétences peuvent être développées et inférées. Comme l'approche par compétences met l'accent sur la

mobilisation des connaissances intégrées pour résoudre des problèmes complexes, l'enseignement qui vise le développement des compétences, doit aussi mettre des étudiants dans des situations où ils doivent résoudre des problèmes en faisant appel à des connaissances de ces trois domaines. Le type de problème qui favoriserait le mieux le développement de compétences à un niveau de formation avancée serait de type mal défini (Huba & Freed, 2000). Un problème est mal défini lorsque l'énoncé du problème, malgré la présence d'un but précis à atteindre, n'explicite ni les critères concrets pour le résoudre, ni les contraintes ou obstacles, ni les données initiales du problème (Tardif, 1997). Huba & Freed (2000) ajoutent qu'un problème est mal défini lorsque plusieurs solutions sont possibles. Scallon (2004) postule que des problèmes mal définis et qui ont un certain réalisme faciliteraient la transférabilité de la capacité de résoudre ces problèmes à des situations similaires rencontrées plus tard.

Tous ces auteurs favorisent d'utiliser des problèmes mal définis pour l'enseignement dans une approche par compétences. En résolvant ces problèmes, l'étudiant fait appel aux connaissances antérieures, cherche à ajouter de nouvelles connaissances selon la nécessité, propose une solution en argumentant et intègre ainsi des connaissances. De cette façon il construit progressivement sa base de connaissances qui l'aidera à résoudre d'autres problèmes. Un exemple de problème mal défini serait : « Un patient se présente avec une douleur thoracique. Interrogez-le et faites le nécessaire pour proposer un plan de traitement pertinent ». Toujours avec cet exemple, la tâche du résident sera de déterminer les circonstances de la douleur, les pathologies qui causent des douleurs thoraciques à l'effort et celles qui causent des douleurs au repos. Il doit aussi déterminer le test prioritaire à faire selon les circonstances et de nouveau déterminer la signification des anomalies pour arriver à un diagnostic logique. En suivant ces démarches, le résident apprendra sur l'angine, l'embolie pulmonaire et bien plus. Il deviendra plus compétent pour résoudre d'autres problèmes de douleur thoracique. Mais,

il demeure à prouver que l'utilisation des problèmes mal définis en éducation favorise une meilleure construction des connaissances.

Il ne suffit pas non plus de confronter des étudiants à des problèmes mal définis et espérer qu'ils développeront leurs compétences en essayant de les résoudre. Des études sur la résolution des problèmes ont démontré qu'il y a une différence dans la façon dont les experts et les novices résolvent des problèmes et cette différence s'expliquerait par l'organisation de la mémoire en réseau sémantique. Tardif (1997) a cité l'exemple de résolution de problèmes physiques par des experts et des novices dans la recherche de Chi (1982). Selon cette étude, lorsque des experts et des novices sont confrontés à des problèmes et doivent les classer en catégories, avec des sous-divisions, la classification par les experts est nettement plus élaborée. L'explication de ce phénomène serait que l'expert a une vaste base de connaissances spécifiques plus développée et mieux organisée. Toujours selon Tardif, il y a actuellement un consensus dans la recherche en psychologie cognitive selon laquelle l'expert résout un problème de façon efficace en raison de l'étendue et de l'organisation de ses connaissances qu'il peut mobiliser facilement selon les besoins.

Comment alors aider les étudiants à construire et développer le réseau sémantique riche nécessaire au développement des compétences ?

Même si les études en psychologie cognitive n'ont pas encore réussi à répondre clairement à cette question, il est probable que le rôle du professeur est d'aider les étudiants à construire et à organiser une base de connaissances riches et spécifiques aux différents types de problèmes. Un des éléments de base est de leur assurer une base de connaissances spécifiques aux problèmes avant de les aider à développer des stratégies générales de résolution de problème. L'enseignant joue donc un rôle direct et important de guide dans la construction des connaissances de son étudiant (Tardif, 1997; Brooks, 1991).

Pour que l'étudiant puisse mobiliser ses connaissances efficacement pour résoudre des problèmes, Regehr a de plus proposé un autre mécanisme selon lequel les connaissances, lorsqu'elles sont appliquées aux contextes, deviennent plus facilement accessibles pour résoudre d'autres problèmes de la même famille (Regehr, 1996). Ces connaissances devenues des connaissances spécifiques facilitent par la suite l'ajout de nouvelles connaissances grâce au processus d'encodage spécifique. Regehr met ainsi un grand accent sur le développement des compétences par l'acquisition de l'expérience.

En résumé, l'enseignement selon l'approche par compétences devrait favoriser la capacité de résolution de problème par l'exposition aux familles de problèmes à résoudre et mettre un accent important sur l'intégration des savoirs aux dimensions de savoir-être et savoir-faire pour résoudre ces problèmes. Pour cela, les étudiants à un niveau de formation avancée devraient être exposés de façon répétitive à des problèmes mal définis appliqués à des contextes réalistes accompagnés d'un processus de rétroaction. Par ce processus de rétroaction, le professeur guidera l'étudiant dans la construction de ses connaissances et des stratégies pour transférer les connaissances d'une situation à une autre.

L'impact de la rétroaction dans le développement des capacités pour résoudre des problèmes a été rapporté par Needham (1991) dans son étude où des sujets qui reçoivent une rétroaction après chaque résolution de problème réussissent à résoudre d'autres problèmes jusque dans 90 % des cas par comparaison à 60 % chez ceux qui n'avaient pas eu de rétroaction.

Cette rétroaction, même si elle est importante, est très dépendante de la compétence des professeurs à donner des rétroactions. Dans l'approche par compétence, à un niveau plus avancé de formation, les étudiants doivent apprendre éventuellement à être moins dépendants de la rétroaction extérieure et doivent développer leur capacité à s'auto-évaluer comme une étape vers le développement de la métacognition. L'auto-

évaluation constitue l'évaluation de l'étudiant de ses propres performances en se basant sur des critères et des niveaux de performance pré-établis (Legendre, 1993; Laurier, 2005). La métacognition, quant à elle, renvoie au processus par lequel l'étudiant devient conscient de son apprentissage et est capable d'agir en cas de difficultés pour y apporter des correctifs (Allal, 1993 & 1999; Scallon, 2004). Selon ces auteurs, l'auto-évaluation permet d'impliquer activement l'étudiant dans son propre apprentissage et l'aide ainsi à développer son autonomie d'apprentissage, une autre dimension de la compétence.

Au cours de tout ce processus de développement de compétences, le souci d'intégrer la dimension affective devrait être omniprésent dans la résolution des problèmes. Peu d'études se sont vraiment penchées sur le problème d'intégration des domaines non-cognitifs au domaine cognitif. En médecine, les études que l'on retrouve à ce sujet ne font souvent que relever, par l'observation, des comportements problématiques tels que des attitudes condescendantes envers les patients (Beaumier, 1992) ou des attitudes arrogantes et irresponsables de certains étudiants (Papadakis, 1999). Mais, peu a été dit sur la façon d'enseigner ou d'évaluer des attitudes désirables pour améliorer la compétence. Huba et Freed (2000) et Wiggins (1998) proposent des tâches complètes lors des activités d'évaluation où la rétroaction doit porter tant sur les facettes cognitive, qu'affective et psychomotrice. Comme nous l'avons précédemment mentionné, dans la formation clinique, les stages cliniques ont été implicitement adoptés comme méthode de formation parce qu'ils permettent de confronter les résidents aux problèmes cliniques véritables qui sont « mal-définis ». Le résident ferait appel à ses connaissances antérieures et pourrait intégrer ses connaissances acquises, ses habiletés affectives et psychomotrices pour résoudre les problèmes.

Ainsi, si un étudiant veut résoudre un problème de douleur thoracique, il aura à faire appel à ses connaissances sur la physiopathologie de l'angine. Il devrait être aussi suffisamment empathique pour réaliser que son patient pourrait craindre d'une possibilité d'angine et pourrait amplifier les symptômes relatifs au système cardiaque et minimiser

des symptômes du système digestif où se situe la source du problème. Le professeur, au cours de la rétroaction, met l'accent sur l'intégration de tous ces domaines et favorise ainsi le développement de la compétence des résidents en douleur thoracique. Chaque problème ainsi résolu, avec une rétroaction de ses professeurs, constituera « une expérience d'apprentissage » enrichissant la base des connaissances de l'étudiant.

III.2 L'évaluation dans l'approche par compétences

En évaluation dans l'approche par compétences, les auteurs tels que Huba et Freed (2000); Scallon (2004); Tardif (1997); Roegiers (2001); Louis et Hensler (2003) et Voorhees (2001) s'entendent pour dire que les compétences ne peuvent qu'être inférées en se basant sur des performances observables quand l'étudiant tente de résoudre des problèmes qui s'apparentent à ceux qu'il devra traiter plus tard en tant que professionnel.

Gagné (2005) insiste pour que l'évaluation soit congruente avec les objectifs de formation, y compris ceux exprimés sous forme de compétences. Selon lui, «... si nous voulons que les étudiants soient des penseurs critiques et des personnes capables de solutionner des problèmes, alors, il faut les confronter avec des situations où ils ont à faire face à ces problèmes ...» (p. 266).

L'une des approches préconisée est l'évaluation dans un contexte authentique. Ce concept a été proposé par Wiggins (1989 et 1998) et repris par plusieurs auteurs dont Huba et Freed (2000) et Scallon (2004). Selon ce concept, l'évaluation des performances se fait avec des problèmes complexes qui surviennent dans un contexte de vie réelle. L'apprentissage et l'évaluation sont de plus en relation étroite, et le processus d'évaluation peut être plus long qu'il n'est habituellement requis pour des tests écrits à réponses ouvertes ou à choix multiples. Les problèmes, selon l'approche par l'évaluation authentique, ont la caractéristique de représenter des problèmes réels que le résident devra affronter plus tard dans la vie professionnelle. Comme il a été déjà mentionné, ils

sont souvent de type « mal défini » avec plusieurs solutions possibles; de telle façon que les stratégies pour résoudre ces problèmes sont également aussi importantes à évaluer que le résultat lui-même (Huba & Freed, 2000). La tâche soumise à l'étudiant fait de plus appel à l'intégration des habiletés (cognitives, psychomotrices et affectives). Rares sont donc les problèmes d'évaluation authentique qui puissent être résolus par de simples rappels des connaissances (Wiggins, 1998).

L'évaluation des tâches dans un contexte authentique oblige souvent à faire appel à un jugement professionnel pour apprécier la qualité du résultat et du processus de résolution du problème. Les auteurs qui favorisent l'évaluation authentique comme approche pour évaluer les compétences favorisent aussi l'utilisation des "rubrics" (ou ce que Scallon (2004) appelle l'échelle descriptive globale). Il s'agit d'une grille où se trouvent la description de la tâche, une échelle de performances avec critères de performances (la description des niveaux acceptable et inacceptable) et les dimensions décomposées de la tâche (Stevens, 2005). Ces auteurs croient de plus que ce type d'échelle descriptive globale favorise le développement de la compétence à l'auto-évaluation. Actuellement, il faudrait des recherches plus poussées pour vérifier ces hypothèses.

III.3 Avantages et limites de l'approche par compétences

À l'heure actuelle, l'approche par compétences permet de répondre aux besoins de souligner l'importance d'intégrer toutes les dimensions du savoir pour résoudre des problèmes parfois très complexes.

Cette approche aurait de plus l'avantage de stimuler et soutenir la motivation dans la formation. L'étudiant apprécie mieux le but et la pertinence des activités d'apprentissage qui lui sont imposées et Tardif (1997) estime que c'est un puissant facteur qui favoriserait l'apprentissage et la construction des nouvelles connaissances.

Un autre avantage de cette approche est la reconnaissance de la complexité du processus cognitif lors de la résolution efficace d'un problème. Cette prise de conscience favorise la planification des activités d'apprentissage plus appropriées et aussi des méthodes d'évaluation formative multidimensionnelles plus adaptées, pour mieux comprendre et mieux aider les résidents lorsqu'ils ont des difficultés.

Mais si l'approche par compétences a des avantages notables, elle a aussi des limites. La première étant actuellement l'absence d'un modèle de taxonomie uniformisée reconnue et acceptée par tous, qui soit comparable à celui de Bloom, modèle de taxonomie la plus reconnue dans l'approche par objectifs, et qui permettrait une meilleure communication entre les responsables de programmes, les professeurs et les étudiants.

L'évaluation dans l'approche par compétence nécessite de plus encore beaucoup de recherches pour déterminer la façon de s'assurer de la validité et de la fidélité des mesures, surtout dans des domaines à performance complexe. À combien de situations complexes doit-on exposer un résident pour inférer sur sa compétence ? Comment vérifie-t-on la fidélité des mesures observées dans un nombre limité de performances complexes. Voilà des questions auxquelles il faudra répondre.

IV Comparaison des deux approches pédagogiques

Le tableau 1 résume les caractéristiques des approches par objectifs et par compétences, ainsi que les avantages et les limites de chaque approche. Le point commun de ces deux approches reste un souci de rendre efficace l'enseignement en tentant d'établir des méthodes d'enseignement et d'évaluation congruentes avec les buts de la formation. C'est surtout par la suite, dans l'application des deux approches dans l'enseignement que la divergence survient. Comme nous pouvons le constater globalement, l'enseignement et l'évaluation selon l'approche par objectifs sont surtout

unidimensionnels, accordant une grande importance à la précision des objectifs opérationnels. L'évaluation ne porte de plus que sur des résultats observables découlant de la performance. Cette approche est centrée sur le professeur et les activités qu'il a rigoureusement prévues pour aider ses étudiants à atteindre les objectifs. Quant à l'approche par compétence, elle insiste sur l'intégration multidimensionnelle et l'appréciation de la qualité du processus de résolution du problème et du résultat qui en découle. L'étudiant y joue le rôle central et il est le principal responsable du développement de ses compétences.

La fidélité des mesures, dans l'approche par objectifs où il y a une atomisation des objectifs, consiste surtout à vérifier la performance de façon répétitive. Mais, selon l'approche par compétences, il est difficile de répéter des performances complexes à plusieurs reprises dans un court laps de temps. La fidélité des mesures devient-elle alors l'analyse des performances dans différentes situations de la même famille de problèmes sur une plus longue durée de temps? Wiggins (1998) et Kane (1999) semblent favoriser cette hypothèse, mais la recherche sur la fidélité des mesures lors des performances complexes est peu abondante et consiste souvent à établir une corrélation interjuge (deux observateurs indépendants au cours d'une performance complexe).

Quant à la validité des tests basés sur les performances complexes où l'on doit inférer la compétence à partir d'un nombre réduit de performances, la controverse concerne surtout la généralisabilité. Aussi, Wiggins (1998) postule que si des situations d'évaluation s'approchent à la réalité, il est possible de concevoir que dans d'autres situations réelles semblables, l'étudiant performerait de la même façon. Kane (1999), favorise plutôt d'échantillonner des situations d'évaluation plus courtes. Messick (1994), recommande d'utiliser des activités d'évaluation basées sur des construits en s'assurant que les situations d'évaluation soient représentatives des domaines de compétences que l'on cherche à évaluer. Cette conception de la validité est différente de la vision

traditionnelle de la validité des tests avec des items unidimensionnels où l'accent est mis sur l'échantillonnage pour s'assurer que des conclusions sur les résultats d'un échantillon de sujets puissent être inférées à la population d'où proviennent ces sujets.

Ce tableau met en évidence un changement de paradigme, où le rôle central du professeur, l'emphase sur les résultats observables et la division des domaines « savoirs, savoir-faire et savoir être » de l'approche par objectifs, laisse place à l'observation à la fois du processus pour atteindre le résultat et du résultat lui-même, l'intégration de connaissances et habiletés psychomotrices et affectives dans la résolution des problèmes et, au rôle central de l'étudiant dans sa propre formation dans l'approche par compétences.

TABLEAU 1 : COMPARAISON DES DEUX APPROCHES; PAR OBJECTIFS ET PAR COMPÉTENCES

	Approche par objectifs (APO)	Approche par compétences (APC)
DÉFINITION ET CARACTÉRISTIQUES	<p>Ce sont des énoncés d'intention qui précisent les changements des comportements découlant de l'enseignement</p> <p>Une distinction entre les domaines de comportements (cognitif, affectif et psychomoteur) grâce à la taxonomie de Bloom</p> <p>Chaque objectif spécifique doit préciser un comportement observable, des conditions de réalisation et des critères d'évaluation</p>	<p>C'est une disposition qui permet la manifestation des capacités stabilisées d'un individu à mobiliser ses connaissances cognitives et habiletés affectives et psychomotrices de façon intégrée pour résoudre efficacement des problèmes complexes dans des familles d'opérateurs.</p> <p>L'énonciation des compétences à développer doit préciser la famille des situations problèmes et clarifier la qualité attendue du produit final</p>
IMPLICATION DANS L'ENSEIGNEMENT	<p>L'APO oblige une planification détaillée des conditions d'apprentissage pour chaque objectif de formation</p>	<p>L'APC implique une exposition aux problèmes à complexité variable, avec une planification des activités de rétroaction selon les critères et standards de réalisation</p>
Rôle du professeur	<p>Son rôle est central et actif dans tout le processus de planification, de prestation et de l'évaluation des savoirs</p>	<p>Son rôle est plutôt le rôle d'un guide qui accompagne l'étudiant dans sa démarche de résolution de problème. De plus, il favorise le développement de la métacognition de l'étudiant par sa rétroaction</p>
Rôle de l'étudiant	<p>Son rôle est plutôt passif : il suit l'ordre des séquences d'objectifs et de cours qui lui est présenté</p>	<p>Son rôle est plus actif : il sera appelé à déterminer ses besoins éducatifs en relation avec des problèmes à résoudre; il construit ainsi activement sa base de connaissances avec l'aide de son professeur</p>
AVANTAGES	<p>L'existence d'une taxonomie universelle permet une structure rigoureuse lors de la rédaction des objectifs</p> <p>La facilitation de la planification de l'enseignement</p> <p>Une vaste base de données sur la fiabilité et de validité des mesures d'évaluation des apprentissages est disponible</p>	<p>Réponse aux besoins d'intégrer les domaines affectifs et psychomoteurs au domaine cognitif pour résoudre des problèmes</p> <p>Le réalisme des activités de résolution de problèmes favorise le maintien de la motivation des étudiants au cours de leur formation</p> <p>La reconnaissance de la complexité du processus de résolution des problèmes permet de mieux cibler l'évaluation formative des apprentissages</p>
LIMITES	<p>La division artificielle des objectifs comportementaux</p> <p>Le degré trop important de spécificité des objectifs rend fastidieux la rédaction des objectifs et leur utilisation</p> <p>La tendance à favoriser des objectifs cognitifs de faible niveau (connaissance, compréhension et application) au lieu du niveau plus avancé (analyse, synthèse, évaluation)</p>	<p>L'absence d'une taxonomie universellement reconnue</p> <p>La complexité des activités de résolution de problème rend difficile la création des problèmes variés et l'exposition des étudiants au nombre élevé de problèmes</p> <p>Manque de données de recherche sur la fidélité des mesures et la validité de l'évaluation des compétences avec des performances complexes</p>

V ENSEIGNEMENT CLINIQUE : QUELLE APPROCHE?

En 1978, Mc Gaghie & coll. ont écrit dans un cahier de Santé publique produit par l'Organisation Mondiale de la Santé que l'enseignement de la médecine devrait favoriser un curriculum axé sur le développement des compétences.

Selon eux, il n'y a pas de compétences universelles puisqu'un médecin compétent devrait être capable d'utiliser une vaste gamme de connaissances, d'attitudes et de comportements pour exercer une activité professionnelle selon les conditions politiques, sociales et économiques des services de santé. Ils ont présenté l'exemple suivant pour illustrer leur prise de position : un médecin diplômé pour exercer dans des pays en développement (comme en Afrique) devrait maîtriser toutes les notions de maladies transmissibles et doit s'attendre à collaborer avec des organismes de santé publique pour s'assurer du développement et du maintien d'environnements salubres; un médecin diplômé en Amérique du Nord devra maîtriser, en plus de la médecine de base, tous les instruments technologiques sophistiqués pour soigner un patient efficacement.

Miller (1990), plusieurs années plus tard, a de nouveau tenté d'explicitier comme suit la notion de compétence en médecine : « ...pour atteindre des objectifs plus larges, les étudiants gradués devraient aussi savoir comment utiliser toutes les connaissances accumulées; sinon ce ne sont que des "savants idiots" ». Il a aussi insisté sur la notion qu'un médecin compétent doit être « fonctionnellement adéquat »; à savoir qu'un gradué doit avoir des habiletés qui lui permettent de trouver l'information à partir d'une variété de sources, d'analyser et d'interpréter des données parfois à première vue disparates, pour en arriver à un diagnostic rationnel et à un plan thérapeutique adéquat.

Au cours de la fin des années 1990, le Collège Royal des médecins et chirurgiens du Canada (1996), a créé un groupe de travail incluant plusieurs organismes provinciaux de santé, d'associations de spécialistes et des facultés de médecine canadiennes, pour produire un document intitulé CanMEDS 2000. Ce document définit les sept compétences fondamentales suivantes qu'un médecin spécialiste doit maîtriser à la fin de sa formation (Frank, 1996) : l'expertise médicale, la communication, la collaboration, la gestion, la promotion de la santé, le maintien de la compétence et le professionnalisme.

Ce document sur les compétences d'un médecin-spécialiste met l'accent sur l'intégration des connaissances (des sciences cliniques, du système de santé et de ses intervenants, des principes déontologiques, des systèmes d'information) des habiletés psychomotrices (pour recueillir des données cliniques ou pour communiquer; que ce soit sur un plan individuel ou en groupe, que ce soit verbal ou par écrit); et aux habiletés affectives (l'acceptation des patients et leurs caractéristiques, la reconnaissance de ses capacités et de ses limites, ainsi que la reconnaissance des rôles de chacun dans une équipe de soins, l'application des valeurs telles que l'intégrité et l'honnêteté dans la prestation des soins médicaux) pour résoudre efficacement des problèmes cliniques.

Le souci central est d'intégrer tous les savoirs pour donner aux patients des soins efficaces et de qualité. La formation clinique par l'approche par compétences implique plusieurs éléments déjà discutés précédemment, dont :

- Le développement des compétences nécessite d'être exposé à des problèmes cliniques variés, mal définis et incomplets. La résolution de chaque problème est une expérience d'apprentissage pour construire progressivement la base des connaissances spécifiques. L'enrichissement et l'organisation de la base de connaissances seront obtenus grâce aux expériences cliniques pratiques variées et

à la rétroaction qui suit toutes les activités d'apprentissage. Ainsi, les discussions de cas théoriques ou réels, la correction des notes consignées au dossier des patients, l'observation directe des performances cliniques avec des patients simulés ou réels sont des exemples d'activités d'apprentissage clinique. Il est important de souligner de nouveau que les compétences seront développées à la suite d'expériences cliniques acquises au cours des stages cliniques puisque ces derniers offrent une grande exposition aux problèmes réels mal définis dans un contexte de pratique proche de celui des futures pratiques de tous les médecins.

- L'évaluation des compétences est faite en observant des performances cliniques avec des problèmes les plus proches possibles des problèmes qui seront éventuellement soumis au résident dans sa vie professionnelle. Des évaluations avec des patients réels ou fictifs, où les résidents doivent réaliser des tâches cliniques complètes, allant du questionnaire jusqu'à la résolution du problème, devraient alors faire partie intégrale de leur formation.

Comme les organismes éducationnels et d'accréditation s'entendent pour donner le rôle central au patient dans des programmes de formation post-doctorale, une réflexion sur la formation clinique des résidents s'impose. Il faut s'interroger sur la façon de mettre le patient au cœur des programmes de formation pour favoriser le développement des compétences cliniques.

Est-ce donc le temps de remettre l'accent sur l'enseignement au chevet des patients ("bedside teaching"); alors que les développements technologiques explosent ? Kirch (1996) a étudié des erreurs de diagnostics dans 400 cas dans un hôpital universitaire et a trouvé que les données cliniques contribuent à obtenir jusqu'à 70 % des diagnostics corrects lorsque l'on fait la corrélation des diagnostics cliniques avec des

résultats d'autopsie. Cette étude renforce la conclusion de deux autres études réalisées antérieurement par Hampton (1975) et Peterson (1992), selon lesquelles 76 % des diagnostics corrects sont portés après un questionnaire clinique et qu'un 10 % additionnel de diagnostics corrects sont portés après l'examen physique. Encore faut-il savoir comment développer ces habiletés cliniques chez nos résidents. Bordage (1999), lors d'une présentation où il s'adressait devant le groupe responsable des affaires éducatives de l'Association américaine de recherche en éducation médicale a tenté de donner des explications cognitives pour des diagnostics manqués. Selon lui, tout repose sur l'acquisition d'habiletés cliniques pertinentes où l'intégration des données cliniques doit avoir une place centrale pour résoudre efficacement les problèmes.

Il va donc sans dire que tout le processus de résolution efficace de problèmes cliniques se base sur les compétences d'effectuer un recueil de données cliniques efficace et pertinent. L'enseignement clinique, s'il veut être congruent avec le but visé, devrait alors prévoir des activités d'enseignement et d'évaluation de ces habiletés cliniques en mettant une emphase sur l'intégration des différents domaines. Un recueil de données cliniques ne peut en effet être efficace et pertinent que si le clinicien sait ce qu'il cherche. En possédant des connaissances sur les maladies, il sait ce qu'il faut chercher; en possédant des habiletés communicationnelles et psychomotrices pour réaliser un questionnaire et un examen physique pertinents, il sait comment aller chercher ces données et augmente ses chances de résoudre le problème du patient. De plus, en créant une relation de confiance qui permet d'obtenir toutes les données psycho-sociales ayant un impact sur la maladie et éventuellement sur le traitement, il serait en meilleure position pour choisir le traitement approprié.

En résumé, pour résoudre un problème clinique, le résident doit aussi être capable d'intégrer des données cliniques, de faire le lien avec les connaissances théoriques et

scientifiques pour arriver à un diagnostic et un plan thérapeutique qui tiennent compte, non seulement des preuves scientifiquement démontrées, mais aussi du contexte social dans lequel il travaille ainsi que des valeurs du patient. Ce type de formation nécessite que le professeur consacre du temps au résident, sur une base individuelle, pour l'aider dans toutes ces démarches cliniques. Il serait utile d'intégrer dans un programme basé sur le développement des compétences cliniques des méthodes d'enseignement structurées et formelles adaptées à ces besoins de formation. L'évaluation formative, par une approche d'évaluation authentique telle que discutée précédemment devrait faire partie intégrante de l'enseignement.

Même si l'approche par objectifs, au cours de cette formation, a un rôle moins important à jouer, elle a quand même son utilité. On peut l'utiliser pour planifier certaines activités ponctuelles théoriques ou pour enseigner certaines habiletés psychomotrices de l'examen physique. Même l'enseignement de certaines habiletés affectives pourrait avoir recours à l'approche par des objectifs (ex.: l'acquisition des connaissances du code déontologique de la profession médicale). Ces activités ponctuelles, unidimensionnelles doivent alors répondre à des besoins éducatifs clairs et explicites.

VI Conclusion

Cet article illustre les avantages et les limites de deux approches pédagogiques que sont celle par objectifs, qui a dominé pendant quarante ans tous les programmes de formation, et celle par compétences qui est en train de prendre une ampleur considérable.

L'approche par objectifs a eu la faveur des professeurs parce qu'elle remettait le centre de l'éducation sur les professeurs et sur leur rôle important dans tout le processus de planification de l'enseignement et d'évaluation à une époque où les besoins de la

société augmentaient, le nombre d'étudiants grandissait et où le souci d'efficacité s'imposait dans l'enseignement.

Mais, depuis 1990, la progression rapide des connaissances dans plusieurs domaines, dont la médecine, force les milieux d'éducation à réfléchir sur les limites de l'approche par objectifs.

L'approche par compétences s'est ainsi imposée puisqu'elle déplace le centre d'attention vers les étudiants et favorise l'implication active de l'étudiant dans son propre apprentissage. Par contre, si le consensus sur la définition de « compétence » émerge, il reste à développer des méthodes d'évaluation des compétences. Ces méthodes devraient être soigneusement étudiées pour s'assurer de la validité des mesures, de la même façon que les méthodes d'évaluation ont été développées dans l'approche par objectifs.

RÉFÉRENCES

Allal, L. (1993). Régulation métacognitive : quelle place pour l'élève dans l'évaluation formative ? In : Allal, L., Bain, D. & Perrenoud, P. (dir.) : *Évaluation formative et didactique du français*, Neuchâtel : Delachaux et Niestlé.

Allal, L. (1999). Impliquer l'apprenant dans le processus d'évaluation : promesses et pièges de l'autoévaluation. In : C. Depover & B. Noël (dir.) : *L'évaluation des compétences et des processus cognitifs; modèles pratiques et contexte*. Bruxelles : De Boeck Université.

Beaumier, A., Bordage G., Saucier D., Turgeon J. (1992). The nature of the clinical difficulties of first year family medicine residents under direct observation. *Canadian Medical Association Journal*, 146 (4) : 489-495.

Ben-David, M. F. (1999). AMEE Guide No 14 : Outcome-based education : part 3- Assessment in outcome-based education. *Medical teacher*, 21(1):23-25.

Bloom, B. S. & coll. (1956). *Taxonomy of educational objectives : the classification of educational goals*. New-York : Mc Kay [cité dans de Landsheere V., 1978]

Bloom, B. S., Hastings J. T., Madaus G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New-York : Mc Graw-Hill.

Bordage, G. (1999). Why did I miss the diagnosis ? Some cognitive explanations and educational implications. *Academic Medicine*., 74 (10) : S138-S143.

Brien, R. (1981). *Design pédagogique : Introduction à l'approche de Gagné et de Briggs*. Québec : les éditions St-Yves.

Brooks, L. R., Norman G. R., Allen S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal. of Experimental Psychology : general*, (120) : 278-287.

Chi, M. T. H., Glaser R., Rees E. (1982). Expertise in problem solving; In : R. J. Sternberg (dir.) *Advances in the psychology of human intelligence*, 161-183. Hillsdale, N. J. : Lawrence Erlbaum Associates.

Cushing, A. (2002). Assessment of non-cognitive factors. In : G. R. Norman, C. P. M. Van der Vleuten, D. I. Newble(dir.) *International Handbook of Research in Medical Education*, 711-755. Great Britain : Kluwer Academic Publishers.

De Landsheere V. & De Landsheere G. (1984). *Définir les objectifs de l'éducation*; 5^e édition. Paris : Presses universitaires de France.

Des Marchais, J. (1996). *Apprendre à devenir médecin, bilan d'un changement pédagogique centré sur l'étudiant*. Sherbrooke, Canada : Presse de l'Université Sherbrooke.

Downing , S.M. (2002). Assessment of knowledge with written test forms. In Norman, G. R., Van Der Vleuten C. P., Newble D. J. : *International handbook of research in medical education*. Great Britain : Kluwer Academic Publishers.

Ewell, P. T. (1985). Assessing Educational Outcomes. *New directions for Institutional Research*, 47. San-Francisco: Jossey-Bass.

Ferland, J. J. (1987). *Les grandes questions de la pédagogie médicale, perspective nord-américaine*; chap. 1 à 5. Laval, Canada : Presses de l'Université Laval.

Flexner, A (1910). *Medical Education in the United States and Canada : a report to the Canergie Foundation for the Advancement of teaching*; bulletin. 4. New-York : The Canergie Foundation.

Frank, K. J. R., Jabbour M., Tugwell P. & al. (1996). Skills for the new millenium : report of the societal needs working group; CanMEDS 2000 Project. *Annals Royal College of physicians and surgeons of Canada*, 29 : 206-216.

Gagné, R. M., Briggs L. J. (1974). *Principles of instructional design*. New-York : Holt, Rinelhart & Winston.

Gagné, R. M. (1985). *The conditions of learning*, (4th ed.). New-York : Holt, Rinehart & Winston.

Gagné, R. M., Wager W. W., Grolas K. C., Leller J. M. (2005). *Principles of instructional design*, (5th ed.). New-York : Thomson Wadsworth.

Garibaldi, R. A., Suhayah R., Moore M., Waxman H. (2002). The in-training examination of Internal Medicine : an analysis of resident performance overtime. *Annals Internal Medicine*, (137) : 505-510.

General Medical Council (1977). Objectives in Medical Education. *Medical Education*, (11) : 241-243.

Hampton, J. R., Harrison M. J. G., Mitchell J. R. A., Prichard J. S., Seymour C. (1975). Relative contributions of history – taking, physical examination and laboratory investigation to diagnosis and management of medical outpatients. *British. Medical Journal*, (2) : 486-489.

Harden, R.M., Crosby J. R, Davis M. H. (1999). AMEE guide no 14 : Outcome-based education, part 1- An introduction to outcome-based education. *MedicalTeacher* : 21 (1) : 7-14.

Harden, R. M. (2002). Learning outcomes & instructional objectives : is there a difference ? *Medical Teacher*, 24 (2) : 151-155.

Holmboe, E. S., Hawkins R. E., Hammett T. W., Mackkrell -Gaglione M. (1996). Current methods in the evaluation of clinical competence. *Society of General Internal Medicine 19th annual meeting*. Washington, D.C., USA.

Huba, M. E. & Freed J. E. (2000). *Learner-centered assessment on college campuses, shifting the focus from teaching to learning* USA : Allyn & Bacon.

Kane, M. (1999). Validating measures of performance. *Educational measurement; issues & practice*, 18(2) : 5-17.

Kirch, W, Schafil C. (1996). Misdiagnosis at a university hospital in 4 medical eras – report on 400 cases. *Medicine*, 75 (11) : 29-40.

Laurier, M.D., Toussignant R., Morissette C. (2005). *Les apprentissages de la mesure et de l'évaluation des apprentissages*, 3e éd. Québec : Gaëtan Morin.

Leach, D. C. (2001). Changing education to improve patient care; *Quality in Health care*, 10 (suppl. II) : ii 54-ii 58.

Le Boterf, G. (1995). *De la compétence : essai sur un attracteur étranger*. Paris : Les Editions d'Organisation.

Le Boterf, G. (2000). *Construire les compétences individuelles et collective*. Paris : Les Editions d'Organisation.

Legendre, R. (1993). *Dictionnaire actuel de l'éducation*; 2e éd. Montréal : Guérin & Paris : ESKA.

Louis, R., Hensler H. (2003). L'évaluation des compétences professionnelles à l'enseignement : perspective nouvelle et implication sur les plans théoriques et pratiques. In Laurier, M. (dir.); *Évaluation & communication, de l'évaluation formative à l'évaluation informative*. Québec : Les éditions Quebecor.

Mager, R. F. (1975). *Preparing objectives for instruction*. Belmont, CA : Fearon.

Mc Avoy (1985). How to choose and use educational objectives. *Medical Teacher*, 7 (1) : 27-35.

Mc Gaghie, W., Miller G., Sajid A., Telder T. (1978). Introduction à un enseignement médical fondé sur l'acquisition des compétences. *Cahier de santé publique*, no. 68, Organisation mondiale de la santé.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2) : 13-23.

Miller, G. (1990). The assessment of clinical skills/competence/ performance. *Academic Medicine*, 65 (9) : 563-567.

Nanson, E. M. (1975). Medical Education 100 years after. *New-Zealand Medical Journal*, (81) : 134-139.

Needham, D. R., Begg I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer : memory-oriented training promotes memory for training. *Memory Cognition*, 1 (19) : 543-547.

Newell, A., Simon H. A. (1972). *Human problem solvin*. Englewood Cliffs, New-Jersey : Prentice-Hall.

Papadakis, M., Osborn E. H., Cooke M., Healy K. (1999). A strategy for detection and evaluation of unprofessional behavior in medical students. *Academic Medicine*, (74) : 980-990.

Perrenoud, P. (1997). *Construire des compétences dès l'école*. Paris : ESF.

Peterson, M. C., Holbrook J. H., Hales D. V., Smith N. L., Staker L.V. (1992). Contributions of the history, physical examination and laboratory investigation in making diagnosis. *Western Journal of Medicine*, 156 : 163-165.

Popham, W. J. (2005). *Classroom assessment : What teachers need to know* (4th ed.). California : Pearson Education.

Pugh, E. W., Lloyd G. J., Mc Intyre N. (1975). Relevance of educational objectives for medical education. *British Medical Journal*, (3) : 688-691.

Regehr, G., Norman G. R. (1996). Issues in cognitive psychology : implications for professional education. *Academic Medicine*, 71(9) : 988-1001.

Reznick, R. K., Blackmore D., Dauphinee W. D., Rothman A. I. & Smee S. (1996). Large scale high stakes testing with OSCE : report from the Medical Council of Canada. *Academic Medicine*, 71(supplement 1) : S19-S21.

Roegiers, X. (2001). *Une pédagogie de l'intégration : compétences et intégration des acquis dans l'enseignement* (2e éd). Bruxelles : De Boeck.

Scallon, G. (1988). *L'évaluation formative des apprentissages, la réflexion (tome 1) et l'instrumentation (tome 2)*. Laval : Presses de l'Université de Laval

Scallon, G. (2000). *L'évaluation formative*. Canada : Editions du Renouveau pédagogique.

Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétence*. Québec : Éditions du Renouveau pédagogique.

Schuwirth, L. W. T., Van der Vleuten C. P. M. (2004). Different written assessment methods : what can be said about their strengths and weaknesses ? *Medical Education*, (38) : 974-979.

Stevens, D. D., Levi A.J. (2005). *Introduction to rubrics : an assessment tool to save grading time, convey effective feedback and promote student learning*. Virginia : Stylus.

Tardif, J. (1993). L'évaluation dans le paradigme constructiviste. In R. Hivon (dir.) : *L'évaluation des apprentissages : réflexions, nouvelles tendances et formation*. Sherbrooke : Editions de CRP.

Tardif, J. (1997). *Pour un enseignement stratégique, l'apport de la psychologie cognitiv*. Québec : Les éditions Logiques.

Tillema, H., Kessels W., Merjers F. (2000). Competencies as a building blocks for intergrating assessment with instruction in vocational education a case from the Netherlands. *Assessment & Evaluation in Higher Education*, 25 (3) : 265-278.

Voorhees, R. A (2001). Competency based learning models : a necessary future. In R.A.Voorhees (dir.) : *Measuring what matters: competency-based learning models in Higher education* . *New directions for institutional research*, (110) : 5-13. San-Francisco : Jossey-Bass.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7) : 41-47.

Wiggins, G. (1998). *Educative assessment : designing assessments to inform and improve student performance*. San-Francisco : Jossey-Bass.

Williams, R. G., Osbourne C. E. (1982). Medical teachers' perspectives on development and use of objectives. *Med. Educ.* (16) : 68-71.

Article 2

**Modélisation d'une grille d'évaluation formative des
performances dans une perspective d'approche par
compétences : un exemple du programme de formation
postdoctorale en Médecine Interne**

Diem-Quyen Nguyen

Département d'administration et fondements de l'éducation
Faculté des sciences de l'éducation

Université de Montréal

© 2006

Résumé

L'approche par compétences en formation universitaire est adoptée de façon élargie, surtout dans des écoles de formation à vocation professionnelle. Les programmes de formation médicale au niveau post-gradué ont aussi emboîté le pas. L'évaluation des compétences devient alors une composante importante du curriculum. Dans ce texte, nous rapportons les démarches de réflexion et de création lors du développement d'un outil de recueil de données qui sera incorporé à une méthode d'évaluation formative des compétences. En premier lieu, nous mettons l'évaluation formative dans le contexte du développement de compétences. Par la suite, nous discutons des différents types d'échelles de mesure, du contexte de formation clinique, de l'analyse des tâches cliniques, de l'établissement des standards et des critères de performance. Finalement, nous rapportons les étapes de la validation du contenu de la grille descriptive globale qui sera utilisée dans la méthode des évaluations formatives de performances cliniques (ÉFPC) par une approche d'évaluation authentique.

Mots clés

Évaluation des compétences – Évaluation formative – Évaluation authentique – Échelles de mesure. -

Abstract

The competency-based approach in higher education is being adopted on a large scale by most of professional training programs. Postgraduate medical education is taking a significant part in this evolution. Assessment of competencies is thereby becoming an area of great importance in such reform. In this article, we will reflect upon formative assessment in the context of development of competencies. We will also report steps leading to the creation of a global descriptive scale to be used in an authentic method of formative assessment of clinical competencies. The discussion will emphasize the roles of different rating scales, analyze the context of clinical training and of the clinical tasks to be assessed, as well as the standard-setting procedure. Finally, the validation of the content on the measurement scale and its results will be reported .

Keywords

Assessment of competencies – Formative assessment – Authentic assessment – Rating scale -

I. L'évaluation formative des compétences

Nos démarches pour créer un modèle de grille d'évaluation formative des performances complexes dans une approche par compétences nous ont amenés à franchir plusieurs étapes de réflexion et de réalisation. En première partie, nous décrivons ces réflexions qui englobent l'évaluation, l'évaluation formative et l'évaluation formative dans une approche de développement de compétences, et dans la deuxième partie, le processus de validation du contenu de cette grille.

I.1 Évaluation formative

Bloom, Madaus & Hasting (1971) définissaient l'évaluation comme une collection systématique d'évidences pour déterminer la quantité et le degré des changements survenus chez l'étudiant après avoir suivi des activités d'enseignement. Cette définition influença longtemps l'évaluation dans l'approche basée sur les objectifs, où seuls des résultats observables sont évalués.

La portée de l'évaluation a été élargie par l'avènement de la pédagogie basée sur les compétences au cours des années 1980. Le milieu pédagogique estime progressivement que le cheminement cognitif de l'étudiant lors de l'accomplissement des tâches complexes serait tout aussi important à évaluer que les résultats. Plusieurs auteurs tels que Gardner (1992), Tardif (1993), Jouquan (2002), Popham (2005) soutiennent que l'évaluation est une activité formelle dont le but principal devrait être celui de fournir à l'étudiant une rétroaction significative. Selon eux, seul un processus d'évaluation qui recueille les résultats et tient aussi compte du processus cognitif pour arriver à ces résultats serait susceptible de donner des informations plus complètes sur les capacités de l'étudiant. Et, au niveau universitaire, l'évaluation fait partie des tâches réalisées par le

professeur pour obtenir des données détaillées et complètes afin d'analyser, juger les performances et discuter avec l'étudiant de ses habiletés et ses capacités (Huba & Freed, 2000).

Trois principales fonctions de l'évaluation des apprentissages sont généralement reconnues : diagnostique, sommative et formative. L'évaluation diagnostique sert à détecter des forces et des faiblesses de l'étudiant, soit pour classer l'étudiant, soit pour réajuster le cours. L'évaluation sommative permet de décider de la promotion, de la réussite ou de l'échec à un cours, un programme ou un stage. L'évaluation formative sert surtout, quant à elle, à suivre la progression de l'étudiant et a lieu au cours de sa formation (Bloom, 1971; de Landsheere, 1984; Harlen, 1997; Scallon, 2000; Popham, 2005).

L'évaluation formative a été formellement définie par Scriven en 1967 alors qu'il abordait le thème de l'évaluation de programme. Ce concept est par la suite repris par Bloom (1971) qui l'a appliqué aux apprentissages scolaires. L'évaluation formative est fondamentalement un processus d'évaluation ayant pour but d'aider l'étudiant à progresser. Selon plusieurs auteurs, anglo-saxons et francophones, différents aspects caractérisent l'évaluation formative quant à ses enjeux, aux rôles de l'étudiant et du professeur et aux implications pédagogiques (De Landsheere, 1984; Scallon, 1988 et 2000, Allal 1991; Albrecht, 1991; Saddler, 1998; Black, 1998; Gipps, 2003). Pour tous, l'évaluation formative devrait avoir lieu lors de la formation, être accompagnée d'un processus de rétroaction et être sous la responsabilité du professeur et de l'étudiant. Scallon (2000); Allal (1991); et Saddler (1998) ajoutent qu'un processus de régulation

vrait aussi faire partie de la définition (la régulation étant, en plus de la rétroaction, une décision-action pour ajuster ou réorienter l'apprentissage).¹

Par contre, l'approche idéale pour améliorer et guider l'apprentissage est encore un sujet de débat. Selon l'approche par objectifs, l'évaluation formative est faite à la fin de chaque unité d'enseignement et, lorsque les objectifs ne sont pas maîtrisés, une reprise de l'unité d'apprentissage est prévue (Bloom, 1971; Scallon, 2000). Dans une perspective de développement des compétences, l'évaluation formative est un processus continu intégré aux activités d'apprentissage où l'observation des performances de l'étudiant au cours des activités d'apprentissage et d'évaluation est un des moyens privilégié pour recueillir les informations pertinentes nécessaires à la rétroaction qui vise à ajuster la formation pour favoriser le développement des compétences non maîtrisées (Scallon, 2004; Huba & Freed, 2000). Dans ces deux approches, la rétroaction est considérée un élément de base de l'évaluation formative. La rétroaction ici est définie comme la communication d'informations à un étudiant suite à une situation d'évaluation (Legendre, 1993).

Les études faites dans différents domaines tels qu'en enseignement secondaire (Butler, 1988), en formation clinique (Hodder, 1989), au niveau des études supérieures (Norton, 2004) et dans la formation des professeurs (Brinko, 1993), reconnaissent l'importance de la rétroaction et son impact positif sur l'apprentissage de l'étudiant. Dans l'approche par compétences, au cours de cette rétroaction, l'étudiant est encouragé à discuter les raisons qui sous-tendent ses performances (Scallon, 2000; Ouellet, 2003;

¹ Laurier et al. (2005) considèrent que le terme « régulation » devrait remplacer le terme « formative » afin de mieux souligner l'importance de la régulation. Cependant, comme le terme « évaluation formative » est utilisé de façon élargie; nous allons utiliser le terme « l'évaluation formative » dans cet article pour désigner le processus d'évaluation qui survient au cours de la formation et où la régulation des apprentissages est une composante essentielle.

Albrecht, 1991; Huba & Freed, 2000). Cette discussion a pour but de mieux comprendre le processus cognitif de l'étudiant pour réajuster les activités

d'apprentissage. Au cours de cette discussion, le professeur tenterait de plus d'identifier ce que l'étudiant pourrait faire différemment pour avoir une meilleure performance (Torrance, 2002).

Le rôle du professeur au cours du processus de rétroaction est celui d'un guide qui s'adapterait à l'étudiant. D'ailleurs, selon certains auteurs, aucun étudiant n'évolue de la même façon qu'un autre de telle sorte qu'une individualisation de l'enseignement selon les besoins éducatifs serait une autre caractéristique de l'évaluation formative (Perrenoud, 1981; Allal, 1981). Dans l'approche par compétences, l'étape de l'individualisation de l'enseignement a lieu lors de la rétroaction après le processus de résolution de problème (Allal, 1991 & Scallon, 2004).

Le souci de développer la métacognition chez l'étudiant est une implication additionnelle de l'évaluation formative selon l'approche par compétences. Cette étape semble être une étape-clé dans le processus du développement de l'autonomie de l'étudiant. À ce sujet, Saddler (1998) écrivait : « ... le but ultime de l'éducation n'est pas seulement de favoriser l'acquisition des connaissances mais aussi de rendre l'étudiant progressivement indépendant du professeur et être capable d'apprendre seul tout au long de sa vie... ». Allal (1999); Gipps (2003); et Scallon (2004) favorisent l'auto-évaluation comme une étape vers le développement de la métacognition. L'auto-évaluation fait allusion au processus au cours duquel l'étudiant évalue sa propre performance en se servant d'un référentiel externe (Allal, 1999). La métacognition réfère à la capacité de l'étudiant de connaître et juger son propre processus d'apprentissage pour s'améliorer (Son & Schwartz, 2002).

En conclusion, soulignons que l'évaluation formative des compétences est un processus dont le point central est l'évaluation des compétences au cours de la formation en mettant l'accent sur l'individualisation de l'enseignement et de l'évaluation. Le développement de l'auto-évaluation et de la métacognition devrait de plus occuper une place importante dans l'évaluation formative pour développer des compétences.

I.2 Évaluation des compétences

Le concept de compétences a, depuis 15 ans, été sujet de plusieurs débats mais un consensus semble actuellement en émerger. La compétence serait une manifestation des capacités d'un individu à mobiliser ses connaissances cognitives, ses habiletés affectives et psychomotrices de façon intégrée, en faisant des choix pour résoudre efficacement des problèmes complexes dans des familles de situations (Voorhees, 2001; Gagné, 2005; Roegiers, 2001; Perrenoud, 1997; Le Boterf, 2000). Les compétences devraient se traduire par des actions efficaces, où l'intégration des connaissances multidimensionnelles est nécessaire pour résoudre des problèmes dans des situations complexes et nouvelles (Louis, 2003).

Mais, si le consensus sur le concept de compétences semble se profiler, il y a par contre peu d'études sur ce que devrait être l'évaluation formative des compétences dans un programme de formation professionnelle universitaire. Quel devrait être l'aspect d'un instrument de mesure des compétences dans une perspective d'évaluation formative ? Comment peut-on s'assurer de la validité et de la fidélité des mesures dans l'évaluation formative des compétences basée sur des performances complexes ?

Les compétences sont inférées des performances complexes pour résoudre des problèmes qui s'apparentent aux problèmes que l'étudiant rencontrera dans sa vie professionnelle (Tardif, 1997; Huba & Freed, 2000; Voorhees, 2001; Louis, 2003; Scallon, 2004; Gagné, 2005). L'évaluation authentique est une approche émergente et qui

prend actuellement une grande importance dans l'évaluation des compétences. C'est un concept proposé par Wiggins (1989), repris et bonifié par d'autres comme Resnick (1992) et Pellegrino (1992). En 1998, Wiggins a clarifié cette approche en proposant plusieurs critères. Ainsi, pour que l'évaluation soit une évaluation authentique, elle devrait inclure des tâches réalistes, une activité qui demande à l'étudiant du jugement et de l'innovation (une simple récitation des connaissances n'est pas qualifiable). Il faudrait aussi que la performance amenant à la résolution du problème dépasse une simple démonstration. Le contexte des performances devrait de plus être proche du milieu de travail éventuel, le problème à résoudre doit faire appel à l'intégration des habiletés et des connaissances et, finalement, on devrait intégrer à l'évaluation un mécanisme de rétroaction pour améliorer la performance. Selon cette approche, la précision des attentes est réalisée avec des échelles de mesures descriptives globales (terme utilisé par Scallon (2004) pour traduire le terme "rubrics" utilisé en anglais) où sont décrits clairement des standards de performances et des standards de contenu (termes utilisés par Messick, 1994). Ce type d'échelle semble, selon certains auteurs, favoriser le développement de l'auto-évaluation, une étape vers le développement de la métacognition (Scallon, 2004; Allal, 1999; Black, 1998; et Gipps, 2003). Pour commencer, nous décrivons les instruments de mesures qui permettent de recueillir les données, une étape essentielle dans toutes les démarches d'évaluation.

I.3 Grille d'évaluation pour des performances

Plusieurs méthodes d'évaluation basées sur la mesure des performances complexes qui permettent l'inférence des compétences ont été rapportées. Glasner (1999) en a décrit quelques exemples dans l'enseignement supérieur : le portfolio (qui est constitué de réalisations de l'étudiant accompagnées de ses réflexions sur sa propre démarche d'apprentissage), le rapport des projets, l'évaluation des performances dans des milieux de travail, les études de cas (où les étudiants appliquent leurs connaissances pour

résoudre des cas rapportés qu'ils seraient susceptibles de rencontrer dans leur vie professionnelle), les journaux de bord de recueils des incidents critiques avec des analyses réflexives et les ECOS (évaluations cliniques structurées et objectives, constituées d'un circuit de 10 à 15 stations d'évaluation d'une durée de 5 à 10 minutes pour vérifier des performances spécifiques, telles que l'examen d'un genou, ou le questionnaire d'un patient ayant une douleur abdominale). Un des principaux outils de cueillette de données utilisés par ces méthodes d'évaluation est constitué de grilles d'évaluation comprenant différents types d'échelles de mesure, qu'il s'agisse de listes de vérification, d'échelles graduées ou d'échelles descriptives (Scallon, 2004; Huba & Freed, 2000; Laurier, 2005).

Ces échelles de mesure sont très populaires en éducation puisqu'elles ont été développées pour contrôler «... l'extrême subjectivité des déclarations non-structurées là où il y a une déficience d'un contenu commun ou d'un standard de référence d'une personne à une autre ... » (Wolf, 1997). Mais, la création des échelles de mesure force à décider du type d'échelle à utiliser. Cela va des échelles avec des choix obligatoires (dont l'exemple fréquemment cité est la liste de vérification avec une échelle binaire), à des échelles de type ordinal utilisant des standards descriptifs. Toujours pour créer une échelle de mesure, le type de catégories de réponses tel que le pourcentage, l'accord et le désaccord, le descriptif ou le comportemental, devrait être clarifié et des valeurs devraient lui être assignées. Il faudrait aussi créer des items unidimensionnels pour représenter un construit (Wolf, 1997). Pour prendre des décisions sur le type d'échelle, le nombre de catégories de réponse, les items qui y figurent, il est d'abord nécessaire de connaître le but pour lequel une grille d'évaluation est créée (Thorndike, 1991; Messick, 1993(a); Wolf, 1997).

Notre recherche vise à créer une grille d'évaluation formative des compétences dans un contexte d'évaluation authentique en nous basant sur des performances complexes. Ce modèle de grille a comme but d'aider l'étudiant à progresser en mettant

l'accent sur l'individualisation de l'enseignement et sur le développement de l'auto-évaluation. Cette grille sera incorporée à une nouvelle méthode d'évaluation formative des compétences cliniques dans la formation médicale au niveau post-gradué du programme de médecine interne que nous appelons l'Évaluation Formative des Performances Cliniques (ÉFPC). Comme la création de la grille d'évaluation oblige une prise de décision quant aux types d'échelles de mesure, regardons ce que sont ces échelles, leurs avantages et leurs limites.

I.4 L'ÉCHELLE DE MESURE

En évaluation basée sur des performances complexes, l'échelle aux choix obligatoires ayant des catégories binaires (par exemple, l'accord ou le désaccord) permet difficilement d'évaluer les compétences. D'ailleurs, si les listes de vérification ont l'avantage d'être complètes et faciles à utiliser (puisque l'on se contente de relever la présence ou l'absence des faits), elles ne permettent pas de détecter le niveau de compétence. Hodges (1999) et de Regehr (1998) ont en effet démontré que la liste de vérification n'a pas réussi à prédire les niveaux de compétence comme a pu le faire l'utilisation de la grille d'évaluation par des experts avec une échelle de mesure ordinale au cours des ECOS en médecine.

Ces échelles de mesure de type ordinal peuvent prendre des formes très variées, allant de l'échelle numérique à l'échelle descriptive (Laurier, 2005). L'avantage de l'échelle ordinale est qu'elle permet d'apprécier la qualité sur un continuum, et, comme l'évaluation des compétences basée sur des performances complexes requière une appréciation qualitative (à la fois du processus cognitif et du résultat), l'utilisation de ce type d'échelle devient attrayante pour évaluer les compétences.

Le principe de construction de ces échelles ordinales est de permettre à différentes personnes de différents points de vue de pouvoir répondre différemment aux items

requérant un jugement qualitatif (ex. : à l'item : « l'étudiant a été attentif au cours de l'entrevue avec le patient », l'échelle sera « 1 - totalement en désaccord; 2 - partiellement en désaccord; 3 - partiellement en accord et 4 - totalement en accord) (selon Likert, 1932; cité par Dunn – Rankin, 1997). De plus, pour des items unidimensionnels, l'échelle de réponse linéaire sur un continuum permet de comparer les items et les personnes (Thurstone, 1925 et 1926, cité par Andrich, 1997b). La prudence s'impose cependant puisque l'assignation à une échelle ordinale des valeurs de distance égale découle des modèles de transformation mathématique selon Andrich (1997a) et Crocker (1986). Ces auteurs rappellent que l'appréciation de la qualité avec ces échelles comporte un jugement subjectif. De cette subjectivité, découle un danger de chevauchement d'un niveau à un autre dans un domaine où interviennent des jugements personnels. Nunnally (1978) et Thorndike (1991) suggèrent de clarifier chaque niveau de performance pour mieux contrôler ce biais.

Dans le contexte de l'évaluation authentique considérée comme une approche pour évaluer des compétences, le modèle d'échelle de mesure recommandé est ce que Scallon (2004) appelle l'« échelle descriptive globale ». Elle intègre ce que Andrich (1997a) appelle l'échelle à standards avec des catégories de réponses sous forme de description d'un standard; catégories qui ont comme point d'ancrage soit des descriptions comportementales ("*behavior anchoring*") ou de performances ("*product anchoring*"). Stevens (2005) présente ce type d'échelle comme un outil de mesure qui décrit la tâche demandée, et qui donne les dimensions de cette tâche. De plus, la différenciation entre les niveaux acceptables et inacceptables est détaillée et spécifiée. Notons que ce type d'échelle diffère de l'échelle descriptive ("*descriptive scale*") telle que l'entend Morgenstern (1997). Les échelles classiquement appelées descriptives contiennent des items constitués de critères unidimensionnels et d'une échelle de mesure où chaque catégorie de réponse est un comportement descriptif. Selon Scallon (2004), pour distinguer l'échelle descriptive de l'échelle descriptive globale, le changement

fondamental est que ces critères descriptifs séparés dans la première sont regroupés pour former une seule suite d'échelons dans la deuxième. Plus d'une vingtaine de livres et de très nombreux sites Web donnent des exemples variés de l'utilisation de l'échelle descriptive globale en éducation. Le point convergent de ces écrits semble être la perception que cette grille permet un meilleur recueil des données pour s'assurer d'une rétroaction de qualité mais peu d'études ont été rapportées pour s'en assurer.

Il n'y a pas non plus véritablement de consensus sur le nombre de catégories qu'une échelle devrait avoir. Selon Andrich (1997a), le nombre de catégories devrait être assez grand pour profiter de la capacité de jugement et de discrimination d'un évaluateur. De façon empirique, selon lui, il est habituel de voir des échelles de quatre à cinq niveaux dans un cas d'échelle unipolaire (allant de « pauvre, adéquat, bon à excellent ») et de cinq à neuf niveaux pour des cas d'échelle bipolaire (du type « fortement en désaccord, désaccord, indécis à d'accord, fortement en accord »). Cet état d'imprécision se reflète aussi dans les études sur l'observation directe et l'évaluation de performances cliniques. Ainsi, dans une perspective d'évaluation sommative, les listes de vérification de type dichotomique « fait ou non fait » ont été utilisées au tout début de la création des ECOS. Progressivement, ces listes de vérification ont fait de la place aux échelles de mesures allant de trois jusqu'à neuf niveaux, avec des niveaux de fiabilité atteignant 0,68 avec 10 sessions d'observation directe (Regehr, 1998; Norcini, 1995 & 2003) pour mieux discriminer les différents niveaux de compétences.

L'évaluation formative des compétences implique à la fois une vérification de la réalisation d'une tâche demandée et une appréciation de la qualité du travail de cette réalisation, puisque la notion de savoir-faire efficace est au centre du concept de compétence. Dans cette perspective, une grille d'évaluation d'évaluation formative des compétences basée sur des performances complexes devrait permettre d'apprécier le processus et les démarches de résolution de problème au cours de la performance. Elle devrait de plus permettre de recueillir des données de qualité adéquate pour fournir une

rétroaction de qualité (Scallon, 2000). Pour atteindre ce but, il a été décidé de créer une grille d'évaluation avec des standards et une échelle ordinale pour apprécier la qualité de performances cliniques. Les items qui s'y retrouvent sont des tâches à accomplir et leurs catégories représentent différents niveaux de standards de performances allant de novice à expert.

II. Contexte et contraintes de l'évaluation formative des compétences cliniques

L'évaluation des performances cliniques dans des situations complexes et variables exige une intégration des savoirs, savoir-faire et savoir être, pour interroger les patients, les examiner, établir une relation thérapeutique et résoudre le problème clinique avec un diagnostic différentiel logique, ainsi qu'un plan d'investigation et de traitement approprié. Tous les aspects de ces compétences cliniques ont été évalués par le passé sous forme de petites unités testant spécifiquement chaque objectif (interroger un patient affecté d'une douleur thoracique, examiner le cœur d'un patient et diagnostiquer à partir d'une vignette clinique) ou par des évaluations globales de stage. Les méthodes d'évaluation de performances disponibles en médecine sont multiples et chacune a des limites et des avantages selon le contexte de formation où serait insérée une méthode d'évaluation formative des compétences cliniques.

La formation médicale en Amérique du Nord et dans plusieurs pays européens est composée de deux grandes parties : la première, communément appelée « la formation de premier cycle » est une formation de quatre à six années et constitue la base commune à tous les étudiants. Elle est composée d'une partie théorique avec des cours en classe et d'une partie pratique avec des stages cliniques dans les hôpitaux. À la fin de la formation, un examen de synthèse facultaire ou national vérifie l'atteinte des objectifs. C'est habituellement un examen théorique avec des questions à choix multiples. Dans certains

pays, comme le Canada et les États-Unis, une composante additionnelle d'examen clinique sous forme d'ECOS est ajoutée pour évaluer des habiletés cliniques de base.

La réussite à cet examen donne le droit de poursuivre la deuxième partie des études médicales de cycles supérieurs. Ces études postdoctorales sont dirigées spécifiquement vers le type de pratique que chaque étudiant a choisi, allant de la médecine familiale générale jusqu'à la médecine spécialisée pointue, comme l'ophtalmologie, en passant par des spécialités générales comme la médecine interne.

Au niveau postdoctoral, la méthode de formation principale est composée de stages cliniques dans les hôpitaux universitaires ou communautaires. Ces stages consistent essentiellement à placer un étudiant dans son futur milieu de travail pour qu'il apprenne à résoudre des problèmes réels à complexité variable, en faisant appel à ses connaissances cognitives, ses habiletés psychomotrices et affectives. La résolution de ces problèmes, sous la supervision d'un médecin compétent du même domaine, permettrait à l'étudiant de bâtir et consolider ses bases de connaissances et d'atteindre éventuellement le niveau de compétence nécessaire pour une pratique médicale indépendante.

Les avantages de la formation par des stages sont multiples et reconnus par plusieurs auteurs tels que Villeneuve, 1994; Toohey, 1996; Upitis, 1999; Cooksey, 1999. L'avantage principal est le développement progressif de compétences professionnelles dans un contexte réel de futur travail professionnel de l'étudiant. Les stages comportent par contre plusieurs contraintes importantes : la capacité d'accueil limitée de chaque service hospitalier qui ne permet pas à tous de faire les mêmes stages en même temps dans le même milieu; chaque étudiant évolue différemment, profite de chaque stage différemment selon ses acquis précédents. De plus, l'aspect aléatoire des problèmes cliniques rencontrés au cours des stages rend aussi difficile la gestion de la planification exacte du contenu du stage. Finalement, la variabilité des compétences pédagogiques des maîtres de stages, souvent choisis pour leur compétence clinique plutôt que pour leur

compétence pédagogique, rend aussi inégal l'apprentissage de chaque étudiant. Malgré ces problèmes, chaque étudiant doit atteindre un niveau de compétences pré-déterminé à la fin de sa formation (Woolliscroft, 2002; Holm, 2002).

La difficulté de planifier des activités d'évaluation formative dans ce contexte a toujours été un défi en formation clinique post-graduée, tel que nous le constatons dans les articles de réflexions de Maudsley, 1989; Borman, 1997; Karle, 1995; Klessig, 2000 et Quirk, 2000. L'importance d'une évaluation efficace au cours d'une formation clinique pour améliorer l'apprentissage des résidents¹ est reconnue par tous pour éviter des échecs à l'examen de certification. Malgré cette constatation, les méthodes d'évaluation actuelles ne semblent pas répondre adéquatement aux besoins de formation des étudiants (Usatine, 1997; Cogbill, 2005; Farrell, 2005). L'évaluation formative au cours de la formation post-graduée reste un domaine qui reçoit peu d'attention en recherche. La plupart des recherches portent plutôt sur des méthodes d'évaluation sommative. Parmi les méthodes d'évaluation des compétences cliniques, l'outil favorisé est la fiche d'évaluation de stage qui reste la plus utilisée, la plus étudiée et la plus critiquée (Turnbull, 1998; Pangaro, 1999; Holmboe, 1998).

La fiche d'évaluation de stage sert à évaluer différents domaines de compétences cliniques dont certains très difficiles à évaluer par d'autres outils (ex. : l'interaction avec l'équipe de travail ou le comportement professionnel). De plus, elle est souple et facile à utiliser. La recherche sur sa fiabilité a démontré qu'un niveau de 0,6 à 0,8 de corrélation inter-juge peut être atteint si elle est utilisée au moins sept fois par année dans une perspective sommative par différents professeurs (Carline, 1992). Par contre, cette fiche d'évaluation de stage est contestée parce que les utilisateurs (notamment les professeurs) ignorent souvent les objectifs de stage, ne connaissent pas suffisamment les étudiants et

¹ Résident : Mot utilisé pour désigner les étudiants de médecine au niveau de formation postdoctorale

finissent par la remplir inadéquatement (Gray, 1996; Turnbull, 2002). Carline (1992) postule que d'autres sources contribuent à ces difficultés : de multiples évaluateurs impliqués, les biais par une connaissance préalable des étudiants, l'observation de différentes performances lors de l'évaluation du même étudiant, et les différents endroits où a lieu l'évaluation. Il en conclut que le contexte de l'utilisation de l'instrument de mesure pourrait affecter la fidélité de cette fiche d'évaluation de stage. La compréhension qu'a l'étudiant de ce qui est attendu de lui en est affectée.

Afin d'améliorer la validité des inférences découlant des performances observées, Turnbull (2000) rapporte un système d'évaluation formative mettant l'accent sur l'échantillonnage des situations cliniques où l'étudiant est évalué par de multiples intervenants (superviseur, patient, infirmière ...) avec une rétroaction orale et écrite. Le système est cependant complexe à mettre en place et la rétroaction à l'étudiant peut être éloignée de la situation d'évaluation. Les données manquent encore pour apprécier la façon dont l'étudiant apprécie ce système et comment il contribue à la progression de l'étudiant. D'autres ont tenté d'améliorer ces évaluations de stage en tentant d'améliorer la fiche d'évaluation. Comme ces fiches d'évaluation de stage utilisent de façon prédominante une échelle de mesure globale de type Likert avec habituellement des descriptifs comme : exceptionnel, excellent, moyen et inacceptable, il y a place à beaucoup d'interprétation. Pangaro (1999) propose d'utiliser des fiches d'évaluation contenant des descriptions qualitatives pour mieux communiquer les forces et les faiblesses de l'étudiant mais, malheureusement, il n'a pas été rapporté comment l'étudiant progresse avec cette méthode.

D'autres méthodes d'évaluation des performances cliniques rapportées sont la revue des dossiers cliniques, l'examen oral en se basant sur des cas cliniques réels ou simulés en posant des questions sur les connaissances, l'examen écrit avec des questions à choix multiples pour évaluer les connaissances cognitives, et l'ECOS pour évaluer des composantes de performance clinique (Holmboe, 1998). La recherche rapportée sur les

ECOS favorise l'utilisation de l'échelle descriptive à multiples échelons de mesure pour évaluer des performances cliniques. Selon Petrusa (2002) et Cushing (2002), des niveaux de corrélation inter-juge de 0,6 à 0,8 peuvent être obtenus mais requièrent des situations d'évaluation multiples (habituellement, le nombre de stations de la plupart des ECOS varie de 11 à 16 et dure entre 3,5 heures et 5 heures). Pour des domaines non-cognitifs, des grilles peuvent parfois contenir jusqu'à 68 items. Les études rapportées sur ces méthodes ont très peu fait état de la progression des résidents. Actuellement, même si les ECOS sont très populaires en médecine, elles restent difficiles à implanter pour l'évaluation formative à cause de leur coût élevé et de la complexité de leur organisation (Mavis, 2002). Ces auteurs reconnaissent de plus qu'une des difficultés majeures de l'évaluation des performances, lorsqu'elles sont complexes, est d'en assurer la validité des inférences et la fidélité des mesures. Récemment, l'équipe de Norcini propose la méthode des mini-exercices cliniques (mini-CEX) pour augmenter la fiabilité et la validité de l'évaluation des performances cliniques des résidents. Selon eux, le manque de fidélité des mesures découle de l'insuffisance de la documentation des performances des résidents. Les professeurs n'ont pas observé les résidents lorsqu'ils les ont évalués sur des dimensions cliniques (Norcini, 1995). Leur méthode consiste à observer différentes composantes d'une entrevue médicale dans différents contextes cliniques pendant de courtes périodes de temps. Les professeurs les évaluent par la suite sur sept dimensions; le questionnaire, l'examen physique, le professionnalisme, le jugement clinique, l'habileté à donner des conseils, l'efficacité et la compétence clinique globale. L'échelle de mesure est de type Likert, ayant 9 catégories allant de "insatisfaisant" à "supérieur" (Norcini, 2003).

Depuis le développement des mini-CEX, plusieurs études ont été publiées à propos de leur fiabilité, leur acceptabilité et leur validité prédictive (Durning, 2002; Farrell, 2005; Boulet, 2002; Norcini, 1997; Holmboe, 2003); aucune recherche a pu

mettre en évidence comment l'étudiant réussit à comprendre ce qui est attendu de lui ni comment il progresse au cours de leur formation.

C'est dans ce contexte et avec ces contraintes que nous avons développé les ÉFPC en tentant de tenir compte de la variabilité du bagage de connaissances de chacun des étudiants, d'une limite du temps disponible des professeurs consacré à l'enseignement (toujours dans le contexte de formation en stages et des compétences finales que chaque étudiant doit développer) .

III. Développement des évaluations formatives des performances cliniques (ÉFPC)

Le développement des évaluations formatives des performances cliniques (ÉFPC) est une méthode structurée et formelle d'évaluation formative selon l'approche de l'évaluation authentique des performances cliniques. Au cours de cette activité, la réalisation d'une tâche complète (allant du questionnaire du patient jusqu'à la résolution du problème), avec un patient réel serait demandée à l'étudiant. Cette tâche complète ciblée sur le problème principal du patient devrait permettre une meilleure compréhension du cheminement de chaque étudiant pour résoudre un problème clinique. On postule de plus que cette méthode permettrait une détection plus précoce de ses difficultés pour une rétroaction plus appropriée.

Les ÉFPC ont été conçues pour évaluer les compétences cliniques suivantes :

- Obtenir des données cliniques efficacement en y intégrant des connaissances théoriques de différentes maladies, des habiletés psychomotrices de l'examen physique et affectives selon les normes de la spécialité.

- Communiquer des informations médicales aux patients en faisant preuve d'une intégration des connaissances théoriques, des données cliniques pertinentes et d'un savoir-faire où l'étudiant s'assure de la compréhension adéquate du patient.

- Communiquer aux collègues son cas clinique en présentant des données cliniques pertinentes et un plan thérapeutique adéquat. L'étudiant doit alors démontrer sa capacité d'intégrer les données cliniques, les connaissances générales scientifiques, les inquiétudes du patient, aux impacts sociaux de la maladie sur la vie du patient dans sa proposition du diagnostic, du plan d'investigation et du plan de traitement.

- Développer un esprit critique de ses propres performances et une pratique réflexive incluant des démarches qu'il doit entreprendre pour s'améliorer.

Chaque session d'ÉFPC dure environ une heure. Pendant une demi-heure, sous observation directe, l'étudiant interroge, examine le patient et répond ensuite à une question du patient. Au cours de la deuxième demi-heure, sans la présence du patient, l'étudiant présente, de façon synthétique, ce qu'il a retenu du problème et propose un diagnostic différentiel, un plan d'investigation ainsi qu'un traitement, lorsque c'est pertinent. Ceci devrait durer dix minutes. Par la suite, commence la rétroaction qui combine l'auto-évaluation et l'évaluation par le professeur. L'étudiant et le professeur rempliront ainsi de façon indépendante une grille d'évaluation de la performance de l'étudiant. La discussion qui suit entre l'étudiant et le professeur porte principalement sur les points divergents de la grille, après comparaison des résultats d'appréciation de l'étudiant de sa propre performance et de ceux du professeur consignés sur cette grille. L'accent est mis sur la compréhension des raisons qui ont amené l'étudiant à performer tel qu'il a performé. Une balise de temps est prévue pour souligner l'importance de l'efficacité dans la résolution des problèmes cliniques tel que le veut la définition des compétences, mais le professeur peut l'ajuster selon les besoins de l'étudiant.

L'observation directe au cours de la performance est le moyen privilégié pour recueillir des informations au cours de l'activité évaluative puisque cette stratégie permet de constater la performance réelle de l'étudiant. Une grille d'évaluation descriptive est créée pour servir de support à l'observation directe et à la rétroaction. Cette grille vise à tenir compte des multiples contraintes inhérentes à la formation par stages qui sont notamment l'évolution variable des résidents au cours de leur formation, le nombre élevé de professeurs impliqués dans la formation clinique rendant difficile tous les processus de standardisation professorale dans des activités évaluatives, et, la difficulté d'inférer les compétences à partir des performances avec un nombre limité de cas cliniques. Ainsi, à la suggestion de Messick (1994), une grille avec des critères génériques a été développée pour favoriser éventuellement l'établissement de la généralisabilité. Ces critères reflètent la qualité des activités professionnelles. Une échelle d'évaluation globale descriptive est favorisée pour apprécier la qualité de la performance tout en tentant de s'assurer de la fidélité de mesure dans le recueil d'informations. Les standards de performances tiendraient compte des compétences devant être développées à la fin de la formation telles qu'elles sont dictées par des organismes académiques et d'accréditation (les Collèges des médecins, les Associations médicales, les Facultés de médecine, etc...).

Cette grille devrait faciliter l'évaluation formative de l'étudiant en le guidant à ajuster ses lectures, ses choix de stages, sa façon d'agir et même sa façon d'être.

Les questions de recherche auxquelles répondra la modélisation de cette grille d'évaluation formative des compétences cliniques sont :

- 1- Quels sont les tâches qui devraient y être?
- 2- Quels sont les critères et les standards de performances qui devraient y être explicités ?

3- Combien de niveaux de performance devraient se trouver dans une grille servant à l'évaluation formative ?

4- Est-ce que cette grille favorise l'auto-évaluation et une rétroaction constructive ?

Nous rapportons ici les étapes que nous avons suivies pour créer cette grille d'observation directe des performances cliniques, ainsi que les résultats préliminaires de son utilisation.

IV Création de la grille

Nous avons suivi plusieurs étapes pour créer la grille qui a servi aux ÉFPC car nous voulons nous assurer de la validité éventuelle des inférences qui en découlent, ainsi que de la fidélité des mesures recueillies.

IV.1 Validité et fidélité des mesures

La validité est un jugement évaluatif intégré du degré selon lequel des évidences empiriques et des rationalisations théoriques peuvent soutenir la justesse et la pertinence avec lesquelles les inférences sont appuyées par des mesures (Messick, 1993b). La validité est considérée comme une propriété fondamentale de tout acte évaluatif puisqu'une évaluation valide donnerait des informations utiles pour guider l'étudiant dans son apprentissage (Popham, 2005; et Mislevy, 2003).

Selon Messick (1993b) et les "Standards for Educational and Psychological Testing" (American Educational Research Association, 1999), la validité devrait être un concept unifié et intégré, regroupant différents aspects distincts tels que le contenu, la structure interne, la substance, l'aspect externe, la généralisabilité et les conséquences.

Ainsi, la validité de contenu réfère-t-elle à l'établissement de la représentativité des items du domaine à être évalué en faisant toujours appel à des jugements d'experts du domaine en évaluation. Quant à l'aspect de structure interne, il s'agit des techniques qui s'assurent que les échelles de mesure contiennent les composantes du domaine que l'on veut mesurer. Cet aspect indique le degré de cohérence entre les items du test et les interprétations qui en découlent. L'aspect de substance, estime Messick qui reprend un concept d'Embretson (1983), fait allusion à la congruence entre la performance demandée et le processus véritable dans lequel l'étudiant s'est engagé. L'aspect externe souligne l'importance des évidences qui permettent d'établir la comparabilité entre une nouvelle méthode d'évaluation et celles existantes (cela regroupe les concepts de la validité concourante et de la validité prédictive des tests). La généralisabilité réfère au degré selon lequel les performances de certains nombres d'étudiants dans certains contextes permet de généraliser à d'autres groupes d'étudiants les performances dans d'autres contextes. Finalement, l'aspect des conséquences du test réfère aux implications sociales de l'interprétation des données et aux actions qui en découlent; cet aspect réfère aux biais systématiques et à la perception de justice d'un test.

La fidélité d'un test est la capacité d'un test de mesurer avec exactitude ce qu'il a à mesurer. (Popham, 2005; American Educational Research Association, 1999). Les indices de fidélité sont établis par différentes techniques telles que la réplicabilité (comprenant les formes équivalentes d'un test, le test-retest, la corrélation interjuge et intra-juge lors des performances répétitives) ou l'étude de la consistance interne (établissant l'homogénéité des items pour mesurer un domaine) (Shea, 2002; Laurier, 2005). Ces indices de fidélité sont influencés par les biais, dont certains biais systématiques (l'appartenance ethnique, l'effet halo, l'effet de contraste et les conditions d'administration d'un test à un groupe par rapport à un autre groupe) pourraient affecter le traitement équitable des étudiants (Popham, 2005; A.E.R.E., 1999; Fleming, 1999).

IV.1.1 Validité et fidélité des mesures pour des performances complexes

À partir de ces concepts, examinons comment nous assurer de la validité et de la fiabilité avec des méthodes d'évaluation de performances complexes.

L'évaluation des performances complexes est un défi pour deux raisons principales. Les performances complexes sont longues à évaluer et, par conséquent, le nombre de situations d'évaluation des performances est restreint par la contrainte du temps disponible (Kane, 1992 & 1999). L'évaluation des performances fait de plus appel aux appréciations de qualité et l'aspect de subjectivité dans les jugements peut alors devenir important avec des erreurs de mesure potentielles (Friedman, 1991; Messick, 1994; Swanson, 1995).

Afin de s'assurer de la validité du contenu lors de l'évaluation des performances, Messick suggère d'utiliser l'analyse des tâches pour établir des standards de contenu (ce que l'étudiant devrait connaître et être capable de faire dans le domaine en question) et des standards de performance (référant au niveau de compétence qu'un étudiant devrait développer). Cette stratégie est aussi suggérée par d'autres auteurs qui ont écrit sur cet aspect de la validité, dont Popham (2005) et Huba & Freed (2000). Cette étape fait appel au jugement des personnes reconnues comme compétentes dans le domaine.

En ce qui concerne l'aspect de la structure interne de la validité, Messick propose des modèles de mesure basés sur l'utilisation des échelles descriptives globales contenant des critères de mesure bien définis. C'est au cours de cette étape que la comparabilité de performances devrait être appliquée non pas entre les individus mais plutôt avec les standards de performance. Les échelles de mesure, les grilles d'évaluation et les listes de vérification font partie des outils de mesure utilisés pour apprécier la qualité de performance.

L'aspect de généralisabilité est l'un des plus difficiles à respecter quand on évalue des compétences basées sur des performances complexes puisque l'évaluation par des performances complexes est faite avec un nombre limité de tâches. On assiste ainsi au choix difficile entre des tests vérifiant en profondeur, mais qui prennent beaucoup de temps et des tests qui couvrent un grand nombre de sujets, mais de façon plus superficielle. Pour résoudre ce problème, l'utilisation d'une description générique des tâches complexes à accomplir, pourrait permettre de généraliser à partir d'une tâche effectuée dans un contexte à une autre tâche faite dans d'autres contextes, estiment Messick (1994) et Friedman (1991). La répétition des tâches à accomplir dans le temps, permet de plus de dégager un profil de performance de l'étudiant qui permettrait d'inférer sur les capacités de l'étudiant à partir de certains nombres de situations évaluées à d'autres situations non évaluées (Messick, 1994 et Wiggins, 1998). Mais, ceci reste actuellement un domaine ayant besoin encore de recherches plus poussées. Pour l'instant en ce qui concerne des tests à enjeu élevé, tels que des examens de certification, Messick (1994) recommande l'échantillonnage des tâches à accomplir.

Quant à l'aspect de substance de la validité, Messick (1994) recommande de nouveau de recourir à l'échantillonnage du domaine des tâches à accomplir tout en s'assurant que ces tâches engagent l'étudiant dans la performance exigée.

Au sujet de l'aspect de la comparabilité avec des critères externes, Messick reste vague et recommande surtout l'argumentation théorique pour comparer les construits de deux tests et leur pertinence dans l'utilisation.

Finalement, Messick estime aussi qu'une attention particulière doit être portée sur des conséquences négatives potentielles d'une sous-représentation des construits.

Quant à la fidélité des mesures lors de l'évaluation des performances complexes, elle reste une question non résolue, mais certaines études suggèrent d'utiliser deux

évaluateurs pour établir la fiabilité inter-juge ou d'utiliser la consistance interne pour s'assurer de l'homogénéité des critères lors de l'évaluation des multiples performances et lorsque l'enjeu des évaluations est élevé (Petrusa, 2002).

Concernant le contrôle des biais, Friedman a identifié deux types d'erreurs : les erreurs aléatoires et les erreurs systématiques. Il propose différentes stratégies pour minimiser ces erreurs. Les erreurs aléatoires pourraient être contrôlées par le recours à l'échantillonnage des situations d'évaluation selon lui. Quant aux erreurs systématiques, il est important de tenir compte du jugement d'experts dans le domaine lors de l'analyse des résultats et de la prise de décision (Friedman, 1991).

IV.2 Étapes de développement d'une grille descriptive globale pour évaluer des performances cliniques.

Les étapes allant de la décision du contenu de la grille jusqu'à sa validation ont eu lieu pour développer une nouvelle grille d'évaluation. Les prochains paragraphes rapportent les étapes suivies au cours du processus de modélisation de cette grille.

IV.2.1 La première étape : Décision sur le contenu de la grille

La recension des publications sur l'observation directe des performances cliniques et l'analyse des tâches ont donné lieu à huit tâches principales qui constituent la base de ce qu'un médecin fait lorsqu'il voit un patient en consultation pour la première fois. Le corpus de publications à ce sujet est très abondant et il n'est pas possible de le citer au complet ici. Un article important est celui où sont explicités des critères d'évaluation proposés par le Collège Royal des médecins du Canada qui a mené une consultation élargie auprès de multiples associations de spécialités médicales et chirurgicales (Frank, 1996). D'autres proviennent des responsables de programme tels que les articles de Jouriles (2002), Holmboe (1998), Petrusa (2002). Ces auteurs reconnaissent

explicitement les huit composantes d'une performance clinique. Il s'agit du questionnaire du patient, des techniques d'entrevue, de l'examen physique, de la communication des informations médicales, de la présentation de cas aux collègues, de la proposition d'un diagnostic différentiel et d'un plan d'investigation et thérapeutique pour résoudre le problème du patient. Ces huit tâches sont les huit items de la grille d'évaluation de notre ÉFPC.

L'échelle de mesure favorisée, si nous restons dans la logique de l'évaluation des compétences, est du type d'échelle ordinale avec des standards descriptifs qui devraient permettre d'apprécier la qualité de performance, du niveau le plus faible (novice) jusqu'au niveau exceptionnel (compétence expert). Pour déterminer le nombre de catégories de niveaux de performance, une révision de la littérature sur des méthodes d'évaluation de performances cliniques et des grilles descriptives a permis de voir qu'aucun consensus n'existe. Par contre, des publications récentes de Norcini (1995 & 2003) et Durning (2002) rapportent l'utilisation d'une échelle à neuf niveaux de type Likert avec les mini-CEX pour des fins sommatives. Ils ont obtenu des indices de fiabilité jusqu'à 0,81 (indice de reproductibilité) et de consistance interne jusqu'à 0,86 après 7 sessions.

L'intégration de ces données a permis de créer dans un premier temps une grille d'appréciation de huit items et neuf niveaux de performance (annexe III), où le caractère « général » de l'échelle laisse beaucoup de place au jugement individuel des professeurs.

IV.2.2 La deuxième étape : Établir la validité du contenu de la grille auprès des experts

Au cours de cette étape, des standards de performance ont été établis. Le processus principal visé ici était d'identifier le degré de performance jugé adéquat pour être reconnu comme compétent.

Plusieurs processus pour établir des standards et des seuils de réussite ont été décrits dans le passé et ont été regroupés en trois catégories par Hambleton (1997) : avec jugement critique sur des données; empirique avec analyse par groupes opposés sur des critères externes; et combiné qui intègre à la fois les principes de la catégorie « avec jugement critique »; et ceux de la catégorie « empirique ». Des méthodes les plus connues ayant besoin des jugements critiques, notons celles qui proviennent d'études sur des tests écrits . Ce sont les méthodes de Nedelsky (1954), d'Ebel (1972) et d'Angoff (1971).

Ces méthodes ont été utilisées principalement pour établir les notes de passage des examens à choix multiples (score de césure). La méthode d'Angoff diffère des deux premières qui obligent d'analyser des choix de réponses (Nedelsky) ou une grille opposant la pertinence et le niveau de difficulté (Ebel). La méthode d'Angoff, elle, demande surtout aux membres du jury d'assigner directement à chaque item la probabilité qu'un étudiant « minimalement compétent » puisse répondre correctement. La moyenne de ces probabilités devient la note de passage de l'item concerné.

L'avantage de la méthode d'Angoff est d'une part qu'elle est simple et d'autre part, que chaque item est traité indépendamment des autres (Erwin, 2001). Des modifications ont été apportées pour l'appliquer à l'évaluation des performances. Mais, les données rapportées sont contradictoires. En médecine, Norcini rapporte plusieurs résultats de recherche où il a utilisé la méthode d'Angoff pour établir des standards de l'examen écrit à choix de réponses à des fins de la certification donnant droit de pratiquer la médecine interne. Selon ses données, les résultats sont comparables à ceux découlant des autres méthodes, restent stables dans le temps avec différents groupes de juges et ne sont pas influencés par le degré d'expertise des juges (Norcini, 1988 et 1992). Par contre, Lin (2005) rapporte qu'en éducation, certains constatent une grande divergence entre l'opinion des juges et l'indice de difficulté des items faciles ou difficiles (Bejar, 1983), et d'autres se demandent si les juges sont véritablement capables d'estimer la capacité d'un étudiant hypothétiquement considéré comme « minimalement compétent » (Shepard,

1993). Finalement, Mitzel (2001) remet en question la pertinence d'avoir des standards variables pour chaque item dans un test et trouve ce processus fastidieux.

Des méthodes combinées ont été utilisées pour évaluer des performances. Il s'agit de la méthode avec un « groupe limite » et la méthode des « groupes contrastes ». Dans le premier cas, des évaluateurs soumettent une liste d'étudiants impossibles à classer « échec » ou « réussite », et leur administrent le test. La médiane de ce groupe donnera la note de passage.

La méthode de groupe contraste consiste surtout à administrer le test à deux groupes d'étudiants que les juges considèrent comme « ayant maîtrisé » ou « n'ayant pas maîtrisé » le domaine évalué. Le point d'intersection de la distribution des scores déterminera la note de passage. Ces deux méthodes reposent de façon importante sur le jugement des évaluateurs pour classer les étudiants.

Une nouvelle méthode émergente proposée par Lewis (1998) est la méthode du signet ("*bookmark standard setting*"). Elle a été proposée pour déterminer les standards de réussite ou d'échec, en déterminant le rang assigné à chaque item (du niveau le plus facile au niveau le plus difficile) en fonction des résultats de performance antérieurs. Selon ce modèle, les juges discutent, en premier lieu, en petits groupes, sur chaque item mesuré et ce qui fait que l'item en question est plus difficile que l'autre. Après cette première discussion, les juges cotent ces items de façon indépendante et il peut en découler un réarrangement des items. Par la suite, au deuxième tour, les juges apprennent la cotation des autres et une discussion a lieu de nouveau sur les items où il y a divergence d'opinion. Après cette discussion, une deuxième cotation sera effectuée pour arriver à un consensus. Au troisième tour, en grand groupe, la médiane des cotations sera donnée à chaque juge. De plus, l'impact sur le nombre d'étudiants qui échoueraient sera estimé et fourni si cette médiane est adoptée comme niveau de note de passage. Finalement, une troisième cotation aura lieu et la médiane des notes obtenues sera utilisée

comme le niveau de note de passage recommandé pour chaque item en utilisant la théorie de réponse aux items, en tenant compte des interactions entre la capacité de l'étudiant, les caractéristiques de l'item tel que son niveau de difficulté et de discrimination.

En médecine, Reznick (1996) puis Dauphinee (1997) ont rapporté des données sur l'utilisation de la méthode Angoff modifiée puis sur la méthode de « groupe limite » au cours de l'examen canadien d'ECOS pour la certification à grande échelle. Avec la méthode Angoff modifiée, la note de passage est établie par la moyenne entre ce que les juges ont déterminé ainsi que le score réel à l'examen, et le taux d'échec s'élevait jusqu'à 17 % dans la première année d'implantation et 13 % au cours de la deuxième année. Lorsque la méthode de « groupe limite » est utilisée avec une échelle de type Likert, le juge, après avoir attribué des notes à l'étudiant selon une liste de vérification, porte un jugement global sur la performance de l'étudiant pour chaque station d'ECOS. La moyenne des scores de tous ceux qui sont considérés comme limite est obtenue et l'ajout d'un standard de déviation a été ajouté et cela constitue la note de passage. Le taux d'échec a ainsi baissé à 9,1 %.

Toutes ces méthodes requièrent des données de performances antérieures qui proviennent des mises à l'essai sous forme de projet pilote pour établir des standards.

Pour notre projet, comme il s'agit d'une nouvelle grille d'évaluation, les professeurs ne disposent pas de données antérieures. La méthode est de plus créée dans un but d'évaluation formative, de telle façon que le but visé n'est pas d'établir une note de passage mais plutôt d'établir les standards considérés comme des niveaux de performance adéquats (ou satisfaisant pour une pratique médicale indépendante) et inadéquats.

Nous nous sommes donc inspirée de la méthode de Lewis pour nous assurer de la validité du contenu des standards de performance.

Au cours de cette phase, vingt professeurs cliniciens ont été sollicités pour agir comme juges du processus d'établissement des standards. Ces professeurs sont recommandés par des membres du comité de programme de médecine interne à l'Université de Montréal. Les critères de choix des juges sont basés sur les recommandations de Stevens (2005), Norcini (2002) et Erwin (2001). Ils suggèrent que les professeurs qui agissent comme juges devraient avoir des connaissances pertinentes du domaine évalué, de la population estudiantine visée, et des méthodes d'évaluation à visée formative. De plus, si ces juges sont de surcroît directement responsables de l'enseignement des étudiants, les impliquer dans le processus d'établissement des standards faciliterait une meilleure collaboration ultérieure pour utiliser cette grille d'évaluation.

De ces vingt professeurs, douze ont accepté l'invitation. Ces douze médecins professeurs proviennent des spécialités médicales diversifiées (neurologie, cardiologie, médecine interne, endocrinologie ...) et ont au moins cinq années d'expérience dans l'enseignement clinique. Ils sont aussi familiarisés avec le programme de formation post-graduée en médecine interne.

L'utilisation de deux enregistrements audiovisuels d'entrevues entre des résidents et des patients vise à faciliter le processus. Ils sont invités à utiliser la grille de l'annexe III pour coter les performances du résident enregistrées sur vidéocassettes. Le but de cette session est d'établir un consensus sur les critères et les niveaux de performance attendus d'un résident.

Le visionnement des enregistrements a eu lieu pendant quatre heures et le processus pour établir des standards s'inspire du modèle de Lewis (1998).

Trois étapes ont été franchies. En premier lieu, ces douze professeurs ont été divisés en trois petits groupes de quatre. Ils ont visionné un premier enregistrement

d'entrevue, et ont coté de façon indépendante, à l'aide de la grille fournie, la performance du résident. Ils ont par la suite comparé leur cotation pour chaque critère et la discussion a eu lieu sur les divergences d'opinion. Un premier consensus du petit groupe a eu lieu sur la cotation de performance du résident. En deuxième lieu, chaque membre du même petit groupe a visionné une deuxième entrevue et a coté de nouveau la performance du résident de façon indépendante et une deuxième discussion a eu lieu car il y avait encore des divergences d'opinion. Par la suite, un deuxième consensus a été atteint sur la cotation donnée pour chaque critère. Cette deuxième étape sert principalement à expliciter et à renforcer les attentes pour chaque niveau de performance. Un porte-parole de chaque petit groupe rapporte par la suite, au cours de la troisième étape où les douze professeurs se sont réunis en grand groupe, la cotation donnée pour chaque critère et une discussion a de nouveau lieu. Ainsi, le résultat final établit les critères des différents niveaux de performance et découle de ces trois étapes de discussion allant de cotation individuelle, à celle en petit groupe puis en grand groupe.

De cette discussion, les conclusions suivantes pourraient être dégagées :

1- Tous s'entendent que les sessions d'observation directe ne doivent pas dépasser une heure et l'accent devrait être mis sur le problème clinique principal du patient (ex. : si un patient se plaint surtout de mal de tête, une douleur abdominale vague ne sera pas abordée) à moins que le résident juge que c'est pertinent. Il lui reviendra par la suite d'expliquer cette pertinence. Le problème en question doit être vague et mal défini, mais réaliste et représenter les problèmes courants en pratique médicale. L'utilisation d'un patient véritable est jugée préférable.

2- La tâche du résident doit englober toutes les démarches cliniques, afin d'arriver à une résolution plausible du problème du patient; les huit items sont donc retenus. De plus, trois professeurs ont tenu à expliciter le questionnaire à propos des impacts de la maladie sur la vie des patients. Cet item a été ajouté comme un neuvième item.

3- Même si tous les professeurs sont d'accord pour utiliser une échelle d'appréciation globale de type Likert, tous se sont rendu compte lors de la discussion, qu'avec une grille d'appréciation globale complètement ouverte, sans description des critères, les attentes de chaque professeur sont très variables. Ils se sont mis d'accord pour mieux expliciter les attentes pour chaque niveau de performance.

4- Ils ont aussi reconnu qu'il leur est difficile de distinguer plusieurs niveaux de performance. Ils suggèrent de retenir trois niveaux: excellent, adéquat et non-acceptable. Comme il s'agit d'une activité d'évaluation formative, ils ne voient pas l'utilité d'une échelle à plusieurs niveaux.

Cette deuxième phase donne lieu à l'élaboration de la grille de l'annexe IV, où nous retrouvons les neuf items et les trois niveaux de performance avec la description de chaque niveau. Les commentaires généraux et les suggestions pédagogiques d'amélioration sont aussi inclus.

IV.2.3 La troisième étape : valider la grille auprès d'un nombre restreint d'utilisateurs

Au début de cette troisième phase, nous avons envoyé la grille de l'annexe IV aux cinq professeurs et aux douze résidents pour nous assurer de la clarté des énoncés, de la compréhension des standards et des critères ainsi que de l'utilisation de cette grille dans le contexte de la méthode d'ÉFPC.

Par la suite, pendant un mois au début de l'année 2005, nous avons mené une étude pilote où les mêmes utilisateurs (cinq professeurs et douze résidents) ont expérimenté la méthode : les professeurs et les étudiants sont assignés l'un à l'autre de façon aléatoire. Une fois assignés, le professeur et l'étudiant fixent la date et l'heure de la

séance d'ÉFPC selon leur convenance. Le professeur prévoit un patient qui lui est connu et qui accepte de participer au projet d'étude sur une base purement volontaire.

Avant la séance, chaque professeur et chaque étudiant reçoit une copie de la description détaillée de la méthode (son déroulement, son but ...) et une copie de la grille d'évaluation. On envoie aussi aux professeurs un texte expliquant comment faire une rétroaction constructive à partir de l'intégration du texte de Brinko (1993) et Ende (1983).

Les patients sont choisis par les professeurs, et un seul critère est retenu; que le patient ait un problème médical véritable courant et stabilisé (afin d'éviter une décompensation clinique au cours de l'AFCP).

Seulement quatre professeurs sur cinq ont pu compléter les séances qui leur ont été assignées, le cinquième a eu un problème de conflit d'horaire (n'étant pas disponible au même moment que le résident). Pour la même raison, neuf résidents ont eu des séances d'ÉFPC.

IV.3 Résultats de la troisième étape

Après un mois d'utilisation, nous avons rencontrés individuellement les professeurs et les étudiants pour recueillir leur perception de la grille d'ÉFPC et de la méthode. Une entrevue semi-structurée a eu lieu. Les questions ont porté sur la compréhension des niveaux de performance, les critères et la perception de l'utilité de la grille pour soutenir une rétroaction.

Des aspects positifs sont regroupés en trois catégories :

- Au sujet des critères : tous se reconnaissent à l'intérieur de cette démarche et sont d'accords que ce sont les neuf étapes de base d'une première entrevue médicale avec un patient. Aucune modification n'a été suggérée.

- Sur le format de l'ÉFPC et de la grille : ils ont aussi été unanimes pour dire que l'ÉFPC avec cette grille permet de centrer l'observation et la rétroaction.
- Quant aux standards de performance, les professeurs trouvent que l'échelle descriptive va dans la bonne direction et les aide en clarifiant mieux les attentes.

Les aspects négatifs concernent surtout l'échelle de standards de performance et la rétroaction :

- De l'échelle des standards de performance :

Les professeurs et les étudiants ont été unanimes quant au nombre de niveaux de performance : une grille avec trois niveaux de performance est trop restrictive. Ainsi, ils ont rapporté que :

... « Certains étudiants performant adéquatement, mais il me semble qu'ils pourraient faire mieux en faisant une certaine manœuvre ... »

... « il a bien questionné, mais il me semble qu'à leur niveau, il aurait dû se renseigner sur l'état d'âme du patient vis-à-vis sa maladie... »

Pour les professeurs, la distinction est parfois floue et ils ont trouvé des chevauchements entre les tâches et les niveaux de performance (ex. : ... [au point 4.1 de la grille de l'annexe IV], « toutes les données sont rapportées, mais c'est trop long, comment faire pour souligner qu'à leur niveau, il faut être plus synthétique? »).

Les quatre ont proposé un niveau additionnel intitulé : « adéquat, mais des améliorations sont nécessaires ».

- De l'auto-évaluation et de la rétroaction :

les étudiants appelés à s'évaluer avec cette grille auraient aimé avoir des exemples pour mieux comprendre les niveaux de performance. De plus, remettre la grille sans explications préalables sème beaucoup de confusion.

Les étudiants insistent pour s'assurer du caractère formatif de l'activité. Ils trouvent que les critères aident à mieux comprendre les attentes mais ne peuvent se prononcer sur l'auto-évaluation, dans le sens d'une meilleure connaissance de soi-même.

V. Discussion et conclusion

Ces étapes de conception, d'établissement des standards de performance et de validation du contenu ont permis de constater que l'évaluation de compétence basée sur des performances complexes requiert un outil de mesure qui soit créé en ayant un but clair de son utilisation. En effet, comme des commentaires des étudiants le soulignent, la grille d'évaluation a été conçue pour des fins formatives et ces étudiants tiennent à cette utilisation de la grille. Il est alors important de s'assurer que si l'on désire la transformer en un outil d'évaluation sommative, d'autres études soient menées.

De plus, nos résultats semblent aussi confirmer qu'une évaluation des compétences requiert une appréciation de la valeur du processus qui mène à la résolution du problème. Pour apprécier et mesurer les performances cliniques, les professeurs reconnaissent l'utilité d'une échelle ordinale avec plusieurs catégories de choix.

Pour établir les niveaux de performance, les trois étapes inspirées de la méthode «onglet» de Lewis ont été nécessaires afin de départager entre ce qui est «acceptable» et ce qui est «inacceptable». Par contre, nous notons que seulement une utilisation par les utilisateurs ciblés permet la détermination du nombre de niveaux de performances. Ainsi, au cours du processus d'établissement des standards, les juges ont trouvé qu'une grille d'évaluation formative des performances complexes ne semble pas en requérir un nombre élevé. Mais c'est seulement suite à l'étape de mise à l'essai avec un groupe restreint de professeurs que nos résultats ont révélé qu'une grille avec quatre niveaux répondrait aux besoins de formation. Par ailleurs, ces catégories bien explicitées semblent faciliter la tâche des professeurs pour observer directement et rétroagir.

Suite à ces étapes, nous avons créé un modèle d'échelle descriptive globale pour évaluer des compétences cliniques. Cette échelle adopte la position où des « sous-activités » peuvent se retrouver à chaque niveau de performance. C'est la grille qui se trouve à l'annexe V. Cette grille respecte le désir des professeurs de maintenir neuf items qui représentent la description de la tâche professionnelle. Comme il a été mentionné, le nombre de niveaux désirés et perçus comme nécessaires pour une grille d'évaluation formative est de quatre, ayant comme type échelle une échelle ordinale avec des descriptifs pour chaque niveau de performance. Le niveau excellent est celui d'un médecin professionnel compétent, le niveau adéquat est celui de performance minimale acceptable. Les deux autres niveaux représentent les niveaux de novices. De plus, l'utilisation d'une échelle descriptive clarifie les attentes et évite des biais potentiels tels que l'effet de halo et l'effet de la complaisance. Par ailleurs, cette grille, en se basant sur des critères génériques, faciliterait éventuellement l'établissement de la comparabilité de performance du résident avec différents cas cliniques. De plus, même avec différents professeurs, le résident peut toujours mieux percevoir ses forces et ses faiblesses pour cibler l'apprentissage afin de s'améliorer, puisque les mêmes critères ont été appliqués.

Le tableau suivant compare cette nouvelle grille utilisée dans des ÉFPC et celle des mini-CEX utilisés dans certains programmes de formation américaine. De nouveau, les différences notables découlent principalement du but pour lequel chaque méthode d'évaluation a été créée.

Tableau de comparaison entre les ÉFPC et les Mini-CEX

	Mini-CEX	ÉFPC
But de l'évaluation	Sommatif	Formatif
Situations de l'évaluation	Segments d'entrevue médicale	Une entrevue clinique entière
Items de la grille	7 items englobant les compétences cliniques et les comportements professionnels	9 items illustrant les tâches professionnelles lors d'une entrevue médicale
Standards de performances	Echelle de Likert à 9 niveaux, les niveaux ne sont pas explicités	Échelle à quatre niveaux avec description de chaque standard de performance
Rétroaction	À la discrétion du professeur	Ciblée sur les points de divergence découlant de la co-évaluation

Notre recherche menée auprès d'un groupe restreint d'utilisateurs ne permet pas d'établir la fidélité des mesures, la généralisabilité à partir d'un nombre restreint de performances. De plus, l'utilité de cette grille, intégrée à une méthode d'évaluation de performance lors d'une utilisation répétée à une population estudiantine élargie, reste aussi à vérifier.

En conclusion, selon nos données, le recueil des données nécessaires pour évaluer des compétences cliniques semble être facilité par l'utilisation d'une grille d'évaluation avec une échelle descriptive globale intégrée à une méthode d'évaluation authentique. Les items qui s'y figurent devraient représenter les tâches professionnelles du domaine évalué; ce qui permet aux utilisateurs de s'identifier facilement. De plus, des niveaux de performance qui traduisent les compétences à développer à la fin de la formation semblent aussi être bien acceptés par des utilisateurs. Ces deux éléments contribuent à accomplir la fonction pour laquelle la grille a été élaborée : faciliter une meilleure communication des informations à l'étudiant pour l'aider à progresser.

RÉFÉRENCES

- Albrecht, R. (1991). *L'évaluation formative, une analyse critique*. Bruxelles : De Boeck.
- Allal, L. (1981). Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application. In Allal, L., Cardinet J., Perrenoud P. (1981). *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.
- Allal, L. (1991). *Vers une pratique de l'évaluation formative; Matériel de formation continue des professeurs*. Belgique : De Boeck & Larcier.
- Allal, L. (1999). Impliquer l'apprenant dans le processus d'évaluation; promesses et pièges de l'auto-évaluation. In C. Depover et B. Noël (dir.) : *L'évaluation des compétences et des processus cognitifs*, pp 35-56. Bruxelles : De Boeck.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington D.C. : American Educational Research Association.
- Andrich, D. (1997a). Rating scale analysis. In J.P. Keeves (dir.), *Educational research, methodology and measurement : an international handbook, (2^e ed.)*, 874-880. Australie : Pergamon.
- Andrich, D. (1997b). Thurstone scale. In J.P. Keeves (dir.) : *Educational research, methodology and measurement : an international handbook (2^e ed.)*. Australie : Pergamon.
- Angoff, W. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (ed.), *Educational measurement, (2^e ed.)*. Washington DC : American Council of Education.

Bejar, I. (1983). Subject matter experts' assessment of item statistics. *Applied Psychological Measurement*, 7 : 303-310.

Black, P., William D. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7-14.

Bloom, B. S., Hasting J. T., Madaus G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New-York : Mc Graw Hill.

Borman, E., Grady P. O. (1997). Postgraduate training policy paper prepared by the subcommittee on post graduate training; Permanent working group of European Junior Hospital Doctors. *Medical Education*., 341 : 3-8.

Boulet, J. R.; McKinley, D. W.; Norcini, J. J.; Whelan, G. P. (2002). Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Advances in Health Sciences Education*, 7 : 85-97.

Brinko, K.T. (1993). The practice of giving feedback to improve teaching : what is effective ? *Journal of Higher Education*, 64(5) : 574-593.

Butler, R. (1988). Enhancing and undermining intrinsic motivation : the effects of task-involving and ego- involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58 :1-14.

Carline, J. D., Paun D. S., Thiede K. W., Ramsey P. G. (1992). Factors affecting the reliability of ratings of students' clinical skills in medicine clerkship. *Journal of General Internal Medicine*, 7 : 506-510.

Cogbill, K. K., O'Sullivan P. S., Clardy J. (2005). Residents' perception of effectiveness of twelve evaluation methods for measuring competency. *Academic Psychiatr*, 29(1) : 76-81.

Cooksey, J., York J. (1999). Lessons learned from the teaching hospitals and the medical education model. *Peabody Journal of Education*, 74 (3 & 4) : 21-32.

Crocker, L. & Algina J. (1986). *Introduction to classical and modern test theory*. California : Wadsworth.

Cushing, A. (2002). Assessment of non-cognitive factors. In G. R. Norman, C. P. M. Van der Vleuten, D. I. Newble (dir.), *International handbook of research in Medical Education*, 711-756. Great Britain : Kluwer Academic Publishers.

Dauphinee, W. D., Blackmore D. E., Smee D., Rothman A. I. & Reznick R. (1997). Using the judgments of physician examiners in setting the standards for a national multi-center high stakes OSCE. *Advances In Health Sciences Education*, (2) : 201-211.

De Landsheere V. & De Landsheere G. (1984). Définir les objectifs de l'éducation, (5^e ed.). Paris : Presses universitaires de France.

Dunn-Rankin, P. & Zhang S. (1997). Scaling methods. In J.P. Keeves (dir.), *Educational research, methodology and measurement : an international handbook (2^e ed.)*, 790-798. Australie : Pergamon.

Durning, S. J., Cation L. J., Markert R. J., Pangaro L. N. (2002). Assessing the reliability and validity of the mini-clinical evaluation exercise for Internal Medicine residency training. *Academic Medicine*, 77 :900-904

Ebel, R. L. (1972). *Essentials of Educational Measurement*. New-Jersey : Prentice-Hall, Englewood Cliffs.

Embretson, S. (1983). Construct validity; Construct representation versus nomothetic span. *Psychological Bulletin*, 93 : 179-197.

Ende, J. (1983). Feedback in clinical medical education. *JAMA*, 250 : 777-781.

Erwin, T. D. & Wise S. (2001). Standard setting. In R.A.Voorhees (Ed.) : *Measuring what matters: competency-based learning models in Higher Education. New directions for institutional research*, (110):55-64.

Farrell, S. E. (2005). Evaluation of student performance : Clinical and professional performance. *Academic Emergency Medicine*, 12(4) : 302.c6-302.c10.

Fleming, N. D. (1999). Biases in marking of student's written work quality. In S. Brown & A. Glassner (dir.), *Assessment matters in Higher Education : choosing and using diverse approaches*. United Kingdom : SRHE & Open University Press.

Frank, K. J. R., Jabbour M., Tugwell P. & al. (1996). Skills for the new millenium : report of the societal needs working group, CanMEDS 2000 Project. *Annals Royal College of physicians and surgeons of Canada*, 29 : 206-216.

Friedman, M. & Mennin S. P. (1991). Rethinking critical issues in performance assesment; *Academic Medicine*, 66 : 390-395.

Gagné, R. M., Wager W. W., Grolas K. C., Leller J. M. (2005). *Priniciples of instructional design*, (5th ed.). New-York : Thomson Wadsworth.

Gardner, H. (1992). Assessment in context : the alternative to standardized testing. In B. R. Gifford & M. C. O'Connor (dir.), *Changing assessments : alternative views of aptitude, achievement and intruction*, 77-121. Boston : Kluwer Academic Publishers.

Gipps, C. & Stobart G. (2003). Alternative assesement. In T. Kellaghan, D. L. Stufflebeam (dir.), *International handbook of educational evaluation*. Boston : Kluwer Academic Publishers

Glasner, A. (1999). Innovations in student assessment : a system wide perspective. In S. Brown & A. Glasner (dir.), *Assessment matters in Higher Education : choosing and using diverse approache*. USA : SRHE & Open University Press.

Gray, J. (1996). Global rating scales in residency education. *Academic Medicine*, 77 (Jan. Sup.), 5 : 55-63.

Hambleton, R. K. (1997). Standard setting in criterion – referenced tests. In P. Keeves (dir.), *Educational research, methodology and measureme : An international handbook*, (2^e ed.). Australie : Pergamon.

Harlen, W & James, M. (1997). Assessment and learning; differences and relationships between formative and summative assessment. *Assessment in Education*, 4 (3) : 365-379.

Hodder, R. V., Rivington R. N., Calcutt L. E. & Hart I. R. (1989). The effectiveness of immediate feedback during the Objective Structured Clinical Examination. *Medical Education*, 23, 184-188.

Hodges, B., Regehr G., Mc Naughton N., Tiberius R., Hanson M. (1999). OSCE checklists do not capture increasing levels of expertise. *Academic Medicine*, 74 : 1129-1134.

Holmboe, E. S., Hawkin, R. E. (1998). Methods for evaluating the clinical competence of residents in Internal medicine : a review. *Annals of Internal Medicine*, 129 : 42-48.

Holmboe, E. S.; Huot, S.; Chung, J.; Norcini, J.; Hawkins, R. E. (2003). Construct validity of the mini Clinical Evaluation Exercise (mini CEX). *Academic Medicine*, 78 (8) : 826-830.

Holm, H. A. (2002). Post-graduate Education. In G. R. Norman, C.P.M. Van der Vleuten, D. I. Newble (dir.), *International handbook of research in medical education*, 381-413. Great Britain : Kluwer Academic Publisher.

Huba, M. E. & Freed J. E. (2000). *Learner-centered assessment on college campuses, shifting the focus from teaching to learning*. USA : Allyn & Bacon.

Jouquan, J. (2002). L'évaluation des apprentissages des étudiants en formation médicale initiale. *Pédagogie Médicale*, 3(1) : 38-52.

Jourriles, N., Burdick N., Hobgood C. (2002). Clinical assessment in emergency medicine. *Academic Emergency Medicine.*, 9 (11) : 1289-1294.

Kane, M. (1992). The assessment of professional competence. *Evaluation & The Health Professions*, 15 (2) : 163-182.

Kane, M. (1999). Validating measures of performance. *Educational measurement, issues & practice*, 18(2) : 5-17.

Karle, H., Nystrup J. (1995). Comprehensive evaluation of specialist training : an alternative board examination in Europe. *Medical Education*, 29 : 38-46.

Klessig, J. M., Wolfsthal. S. D. & Levine M. A., Stickley W., Bing-You R., Lansdale T. F., Battinelli D. L. (2000). A pilot survey study to define quality in residency education. *Academic Medicine.*, 75 (1) : 71-73.

Laurier, M. D., Toussignant R., Morissette D. (2005). Les principes de la mesure et de l'évaluation des apprentissages, (3e ed.). Québec : Gaëtan Morin.

Le Boterf, G. (2000). *Construire les compétences individuelles et collectives*. Paris : Les Editions d'Organisation.

Lewis, D. M. & coll. (1998). The Bookmark standards setting procedure : methodology and recent implementation. Cité par Erwin T. D. & Wise S. L. (2001) Standard setting; In R Voorhees (dir.) : *Measuring what matters : competency-based learning models in higher education : New Direction For Institutional research*, 110 : 55-64.

Lin, J. (2005). The bookmark standard setting procedure; strength and weaknesses. Site web:

http://www.educational.valberta.ca/educ/psych/crame/files/standard_setting.pdf,

Consulté : février 2006.

Louis, R., Hensler H. (2003). L'évaluation des compétences professionnelles à l'enseignement : perspective nouvelle et implication sur les plans théorique et pratique. In Laurier, M. (dir.), *Évaluation & communication, de l'évaluation formative à l'évaluation informative*. Québec : Les éditions Quebecor.

Maudsley, R. F. (1989). Effective in-training evaluation. *Medical Teacher*, 11 (3/4) : 285-290.

Mavis, B. E., Henry R. C. (2002). Between a rock and a hard place : finding a place for the OSCE in medical Education. *Medical Education*, 36 : 408-409.

Messick, S. (1993a). Foundations of validity : meaning and consequences in psychological assessment. Research report; *Educational Testing Service*. New Jersey : Princeton.

Messick, S. (1993b). Validity. In R. L. Linn, *Educational measurement*, 13-104, (3^e ed.). New-York : American Council on Education & Oryx Press.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2) : 13-23.

Mislevy, R. J., Wilson M. R., Ercikan K., Chudowsky N. (2003). Psychometric principles in student assessment. In T. Kellaghan, D., Stufflebeam, L. A. Wingate (dir.). *International handbook of Educational Evaluation*, 489-532. United Kingdom : Kluwer Academic Publishers.

Mitzel, H. C. , Lewis D. M., Patz R. J. & Green D. R. (2001). The bookmark procedure : psychological perspectives. In G. J. Cizek (dir.), *Setting Performance Standards : concept methods and perspectives*, 249-281. New Jersey : Erlbaum.

Morgenstern, C. & Keeves J. P. (1997). Descriptive scales. In J. P. Keeves (dir.), *Educational Research, Methodology and measurement : an international handbook* (2^e ed.), 900-908. United Kingdom : Elsevier Science Ltd.

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and psychological Measurement*, 14 : 3-19.

Norcini, J. J., Shea J. A. & Kanya D. T. (1988). The effect of various factors on standard-setting. *Journal of Educational measurement*, 25 : 57-65.

Norcini, J. J. & Shea J. A. (1992). The reproductibility of standards over groups and occasions. *Applied Measurement in Education*; 5, 63-72

Norcini, J. J., Blank L. L., Aznold J. K., Kimball H. R. (1995). The mini-CEX (Clinical evaluation exercise) : a preliminary investigation. *Annals of Internal Medicine*, 123 : 795-799.

Norcini, J. J.; Blank, L. L.; Aznold, G. K.; Kimball, H. R. (1997). Examiner differences in the mini-CEX. In A. J. J. A. Scherpbier; C. P. M. Van der Vleuten; J. J. Rethans, A. F. W. Van der Steeg (eds) : *Advances in medical education*, 170-172. Dordrecht : Kluwer Academic Publishers.

Norcini, J. J.; Guille R. (2002). Combining tests and setting standards. In G. R. Norman, C. P. M. Vander Vleuten, D. I. Newble (dir.), *International handbook of research in Medical Education*, 811-834. United Kingdom : Kluwer Academic publishers.

Norcini, J. J., Blank L. L., Duffy S., Fortna G. (2003). The mini-CEX : A method for assessing clinical skills. *Annals of Internal Medicine*, 138 : 476-481.

Norton, L. (2004). Using assessment criteria as learning criteria : a case study in psychology. *Assesment & Evaluation in Higher Education*, 29, 6, 687-702.

Nunnaly, J. C. (1978). Psychometric theory, (2^e ed.), chapitre 2 : Scaling models. New-York : Mc Graw-Hill.

Ouellet, A. (2003). Pour une évaluation des apprentissages en lien avec la compétence. In M. D. Laurier (dir.), *Evaluation et communication : de l'évaluation formative à l'évaluation informative*, 59-99. Québec : Les éditions Québecor.

Pangaro, L. (1999). A new vocabulary and other innovations for improving descriptive in-training evaluations. *Academic Medicine*, 74 (11) : 1203-1207.

Pellegrino, J. W. (1992). Understanding what we measure and measuring what we understand. In B. R. Gifford & M. C. O'Connor (dir.), *Changing assessments : alternative views of aptitude, achievement and instruction*, 275-301. Boston : Kluwer Academic Publishers.

Perrenoud, P. (1997). *Construire des compétences dès l'école*. Paris : ESF.

Perrenoud, P. (1981). Les différences culturelles aux inégalités scolaires : l'évaluation et la norme dans un enseignement indifférencié. In Allal, L. Cardinet J., Perrenoud P. (dir.), *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.

Petrusa, E. R. (2002). Clinical performance assessments. In G. R. Norman, C. P. M. Vander Vleuten, D. I. Newble (dir.), *International handbook of research in Medical Education*, 673-710. United Kingdom : Kluwer Academic Publishers.

Popham, W. J. (2005). *Classroom assessment : What teachers need to know*, (4e ed.). California : Pearson Education.

Quirk, M., Stone S., Chuman A., Devaney-O'Neils S., Mazor K., Starr S., Lasser D. (2002). Using differences between perceptions of importance and competence to identify teaching needs of primary care preceptors. *Teaching & Learning in Medicine*, 14 (3) : 157-163.

Regehr, G., MacRae H., Reznick, R. K., Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine*, 73 : 993-999.

Resnick, L .B. & Resnick D. L. (1992). Assessing the thinking curriculum : new tools for educational reform. In B. R. Gifford & M. C. O'Connor (dir.) *Changing assessments : alternative views of aptitude, achievement and instruction*, 37-77. Boston : Kluwer Academic Publishers.

Reznick, R. K., Blackmore D., Dauphinee W. D., Rothman A. I. & Smee S. (1996). Large scale high stakes testing with OSCE : report from the Medical Council of Canada. *Academic Medicine*, 71(1) : 519-521.

Roegiers, X. (2001). *Une pédagogie de l'intégration : compétences et intégration des acquis dans l'enseignement*, (2e ed.). Bruxelles : De Boeck.

Saddler, R. (1998). Formative assessment : revisiting the territory. *Assessment in education*, 5 (1), 77-84.

Scallon, G. (1988). *L'évaluation formative des apprentissages, la réflexion (tome 1) et l'instrumentation (tome 2)*. Canada : Presses de l'Université de Laval.

Scallon, G. (2000). *L'évaluation formative*. Canada : Éditions du Renouveau pédagogique.

Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Québec : Éditions du Renouveau pédagogique.

Scriven, M. (1967). The methodology of evaluation. In *Perspectives of curriculum evaluation; AERE monographs series on Curriculum Evaluation*, 1, 39-83. Chicago : Rand McNally.

Shea, J., Fortna G. S. (2002). Psychometric methods. In G. R. Norman, C. P. M. Van der Vleuten, D. I. Newble (dir.), *International Hadbook of Research in Medical Education*, 97-126. Boston : Kluwer Academic Publishers.

Shepard, L., Glaser R. & Bohrnsteoft G. (ed.) (1993). *Setting performance standards for student achievement*. California : National Academy of Education.

Son, L. K., Schwartz B. L. (2002). The relationship between metacognitive monitoring and control. In T. J. Perfect & B. L. Schwartz (dir.), *Applied Metacognition*, 15-39. United Kingdom : Cambridge University Press.

Stevens, D. D.; Levi A. J. (2005). Introduction to rubrics, an assessment tool to save grading time, convey effective feedback and promote student learning. Virginia : Stylus.

Swanson, D. B., Norman G. R., Linn R. L. (1995). Performance-based assessment : lessons from the Health Professions. *Educational Researcher*, 24 (5), 5-11, 35.

Tardif, J. (1993). L'évaluation dans le paradigme constructiviste. In R. Hivon (dir.) : *L'évaluation des apprentissages : réflexions, nouvelles tendances et formation*. Sherbrooke : Editions de CRP.

Tardif, J. (1997). *Pour un enseignement stratégique, l'apport de la psychologie cognitive*. Québec : Les éditions Logiques.

Thorndike, R. M., Cunningham G. K., Thorndike R. L., Hagen E. F. (1991). *Measurement and evaluation in psychology and education*; (5th ed.). New-York : Mac Millan Publishing company.

Toohy, S.; Ryan G.; & Hughes C. (1996). Assessing the practicum. *Assessment. & Evaluation. in Higher Education*, 21 (3) : 215-227.

Torrance, H., Pryor J. (2002). *Investigating formative assessment : teaching, learning and assessment in the classroom*. Maidenhead : Open University Press.

Turnbull, J., Gray J., MacFadyen J. (1998). Improving in-training evaluation programs. *Journal of General Internal Medicine*, 13 : 316-328.

Turnbull, J. , Gray J., MacFadyen J., Van Barneveld C., Norman G. (2000). Clinical work sampling : a new approach to the problem of in-training evaluation. *Journal of General Internal Medicine*, 15 : 556-561.

Turnbull, J., Van Barneveld C. (2002). Assessment of clinical performance; in-training evaluation. In G. R. Norman, CPM Van der Vleuten, D. I. Newble (dir.) : *International Handbook of Research in Medical Education*. Boston : Kluwer Academic Publishers, 793-810.

Upitis, R. (1999). Teacher education reform : putting experience first. *Teacher Education Quarterly*, 26 (2) : 11-19.

Usatine, R. P, Edelstein R.A., Yajima A., Slavin S. J., Wilkes M. S. (1997). Medical student perceptions of the accuracy of various new clinical evaluation methods. In A. J. J. Scherpbier, C. P. M., Van der Vleuten J. J., Rethans A. F., Van der Steeg (dir.), *Advances in medical education*, 200-202. Boston : Kluwer Academic Publishers.

Villeneuve, L. (1994). *L'encadrement du stage supervisé*. Québec : édition St-Martin.

Voorhees, R. A (2001). Competency based learning models : a necessary future. In R.A.Voorhees (dir.), *Measuring what matters: competency-based learning models in higher education*. *New directions for institutional research*, (110):5-13. San-Francisco : Jossey-Bass.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7) : 41-47.

Wiggins, G. (1998). *Educative assessment : designing assessments to inform and improve student performance*. San-Francisco : Jossey-Bass.

Wolf, R. M. (1997). Rating scales. In J. P. Keeves (dir.), *Educational research, methodology and measurement : an international handbook*, (2nd ed.), 959-965. United Kingdom : Elsevier Science Ltd.

Woolliscroft, J. O. (2002). Medical student clinical education. In G. R. Norman, C.P.M. Van der Vleuten, DI Newble (dir.); *International handbook of research in Medical Education*, 365-380. Boston : Kluwer Academic Publishers.

Article 3

**L'évaluation formative par une approche d'évaluation
authentique :**

**Résultats des Évaluations Formatives des Performances
Cliniques du programme de Médecine Interne**

Diem-Quyen Nguyen

Département d'administration et fondements de l'éducation
Faculté des sciences de l'éducation

Université de Montréal

© 2006

RÉSUMÉ

L'évaluation formative des compétences est un domaine où peu de recherches ont été effectuées pour vérifier quelle méthode d'apprentissage permet le développement des compétences aussi complexes que des compétences cliniques.

Les évaluations formatives de performances cliniques (ÉFPC) ont été développées en appliquant l'approche de l'évaluation authentique à l'évaluation basée sur des performances complexes en y intégrant l'évaluation des professeurs et l'auto-évaluation et en donnant une place prédominante à la rétroaction interactive.

Une cohorte de 13 résidents a été suivie longitudinalement sur six mois. Chacun des résidents a une session d'ÉFPC par mois.

Les résultats montrent une amélioration significative de la performance clinique globale qui est la somme de neuf sous-compétences. Six sous-compétences ont progressé significativement et les trois autres ont aussi progressé, même si elles n'ont pas atteint un seuil de différence statistique significatif.

L'analyse qualitative montre que les six sous-compétences où il y a une amélioration ont été des domaines principaux de rétroaction.

L'entrevue avec des professeurs montre un grand enthousiasme pour cette méthode considérée comme valide pour évaluer les compétences cliniques.

Ces résultats montrent que l'évaluation formative par l'approche authentique, où la rétroaction tient une place importante, favorise le développement des compétences cliniques de façon significative sur une courte période de temps.

Mots clés

Évaluation formative. Évaluation des compétences. Évaluation basée sur des performances complexes. Rétroaction.

ABSTRACT

In the area of formative assessment of competencies it has hitherto been inadequately documented towards the helping of residents to develop complex clinical competencies.

Formative Clinical Performance Appreciations (Form-CPA) have been developed by applying the principles of authentic assessment in performance-based assessment. Furthermore, the teacher's assessments and the residents' self-assessments were integrated in the Form-CPA. Feedback and learning regulation are two of the main components of this method.

A cohort of 13 residents has one Form-CPA per month. The result analysis after six months of follow-up shows a significant change in overall clinical competence, this being the sum of nine clinical sub-competencies. Of those nine, six achieved a significant improvement and the other three show a small improvement.

Qualitative analysis shows that the six areas of improvement were indeed representative of the main area of concerns and feedback from the teachers.

Teacher's interview revealed a high acceptance of the Form-CPA as a valid method to assess clinical competence.

Keywords

Formative assessment – Competency assessment – Performance-based assessment
– Feedback.

I. Introduction

Le concept d'évaluation formative en éducation a été formellement défini par Bloom (1971) et par la suite adopté de façon élargie par le milieu de l'éducation au cours des années 1970. Il s'agit principalement d'un processus ayant pour but d'aider l'étudiant à progresser.

Appliqué au développement de compétences, l'évaluation formative évalue au cours de la formation les manifestations des capacités d'un étudiant à mobiliser ses connaissances cognitives, ses habiletés affectives et psychomotrices de façon intégrée en faisant des choix pour résoudre efficacement des problèmes complexes dans des familles de situations (Nguyen, 2006a).

Dans le programme postdoctoral de médecine, les compétences cliniques comprennent l'intégration des savoirs, savoir-faire et savoir être pour mener efficacement un questionnaire clinique et un examen physique, discuter des informations cliniques avec les patients, présenter le problème clinique à d'autres collègues et proposer un diagnostic différentiel, un plan d'investigation et un plan thérapeutique adéquat (Frank, 1996).

L'évaluation formative des compétences cliniques devrait donc viser à aider les résidents dans le développement de ces compétences. L'évaluation formative des compétences cliniques au cours des programmes postdoctoraux a reçu cependant peu d'attention en recherche, le domaine de l'évaluation sommative des performances cliniques ayant toujours occupé une position prédominante (Yorke, 1998).

L'une des approches émergentes qui favorise le développement des compétences avec la résolution des tâches complexes est celle de l'évaluation authentique. Cette approche insiste sur l'utilisation de tâches complexes et réalistes, en lien direct avec des

tâches professionnelles que l'étudiant devra accomplir dans sa pratique et qui exige du jugement et de l'innovation lors de la résolution du problème dépassant une simple démonstration (Wiggins, 1998). Dans cette perspective, la formation clinique adopte cette approche depuis de nombreuses années. Les outils pour récolter des données en vue d'une évaluation formative par une approche authentique sont par contre moins bien documentés.

La fiche d'évaluation de stage reste l'outil le plus utilisé parmi des outils d'évaluation formative formels rapportés (Turnbull, 1998; Pangaro, 1999; Holmboe, 1998). Elle permet d'évaluer l'atteinte des objectifs dans différents domaines, allant des connaissances jusqu'aux comportements professionnels tels que la collaboration et l'auto-apprentissage. Par contre, cette fiche a été sujette à plusieurs critiques. Puisqu'elle est parfois remplie par des superviseurs de stage qui connaissent peu l'étudiant, on lui reproche principalement, son manque de fiabilité (Gray, 1996). Pour rendre les résultats de l'évaluation plus fiables dans un contexte sommatif, certains auteurs proposent d'utiliser plusieurs prises de mesures par plusieurs évaluateurs pour atteindre un niveau de fiabilité acceptable, tandis que d'autres proposent d'utiliser de multiples évaluateurs en incluant des patients et l'équipe paramédicale pour évaluer différents aspects de la performance des résidents au cours des stages comprenant une rétroaction orale ou écrite (Turnbull, 2000). Cette méthode d'évaluation est appelée « instrument d'évaluation 360° ». La complexité de ce processus rend actuellement cette méthode difficile à implanter. Pangaro (1999) a plutôt proposé d'utiliser des fiches d'évaluation de stage avec des descripteurs qualitatifs par différents superviseurs de stages. En 2002, son équipe rapportait des données sur l'acceptabilité de ces fiches par des étudiants où jusqu'à 50 % des étudiants les trouvaient utiles, mais, à l'heure actuelle, aucune donnée n'existe quant à la documentation de la progression de l'étudiant (Battistone, 2002).

D'autres outils pour l'évaluation en formation clinique comprennent des examens écrits avec des questions à choix multiples, des examens oraux avec de véritables patients ou des patients fictifs. Ces examens vérifient surtout les connaissances théoriques des étudiants (Holmboe, 1998). Norcini (1995; 2003) propose les mini-exercices d'évaluation clinique (mini-CEX) pour observer des étudiants lors de brèves rencontres avec des patients, soit pour questionner, examiner ou soit pour donner des conseils médicaux. Ces auteurs espèrent ainsi rendre des données consignées sur les fiches d'évaluation de stages plus fiables et augmenter la validité des évaluations. À l'heure actuelle, des données manquent encore quant à l'utilité de ces mini-CEX pour favoriser le développement des compétences cliniques. Wakefield (1985), Day (1990), et Holmboe (2004) encouragent aussi l'utilisation systématique de l'observation directe dans l'évaluation des performances cliniques pour valider des mesures consignées sur la fiche d'évaluation de stage. Ils reconnaissent, cependant, des contraintes importantes qui limitent l'évaluation adéquate des performances cliniques.

Une première contrainte de la formation clinique sous forme de stages est la variabilité sur le plan des compétences pédagogiques des professeurs de cliniques, souvent choisis pour leur expertise clinique plutôt que pour des compétences pédagogiques (Holm, 2002).

Une deuxième contrainte peu rapportée est celle de la variabilité dans le bagage de connaissances des étudiants qui évoluent rarement au même rythme, à cause du cheminement varié inhérent à une formation entièrement composée de stages : peu de résidents font les mêmes stages dans les mêmes milieux et verront les même types de pathologies.

Une troisième contrainte qui préoccupe bien des directeurs de programme est celle du temps professoral. Des professeurs cliniciens ont plusieurs tâches à accomplir : travail clinique, recherche, gestion administrative et enseignement (Ferland, 1987; Holm,

2002). Il en résulte qu'une formation professorale approfondie et élargie pour standardiser l'évaluation est considérée comme difficilement réalisable (Jolly, 2002).

Notre recherche vise à vérifier si une méthode d'évaluation authentique comme les évaluations formatives des performances cliniques (ÉFPC) utilisée dans une perspective d'évaluation formative permet de développer des compétences professionnelles au cours d'une formation composée principalement de stages malgré les contraintes discutées ci-haut. Les sous-questions suivantes sont abordées :

1. Est-ce que les ÉFPC permettent d'évaluer la progression de l'étudiant à savoir :
 - 1.1 Quelles sont les compétences maîtrisées par les résidents au cours de la recherche ?
 - 1.2 Quel est le profil de progression des résidents au cours de la même période ?
2. Y-a-t-il une relation entre les compétences maîtrisées et les conseils pédagogiques annotés par les professeurs ?
3. Quelles sont les perceptions des professeurs de l'utilité de cette méthode pour l'évaluation formative des compétences des résidents ?

II. Méthodologie

II.1 Description de la méthode « Évaluation Formative des Performances Cliniques » (ÉFPC)

Le projet des ÉFPC vise à développer une méthode d'évaluation formative formelle et structurée sous une approche authentique pour évaluer des résidents en médecine interne.

Chaque ÉFPC consiste en une séance d'évaluation globale au cours de laquelle les premières trente minutes sont consacrées à l'observation directe d'un résident qui interroge un véritable patient sur son problème médical principal, l'examine par la suite de façon pertinente au problème que le patient lui a rapporté et répond à ses questions . Au cours des 30 minutes qui suivent, en l'absence du patient, le résident présente de façon synthétique les informations au professeur comme il les aurait présentées à un collègue, puis propose un diagnostic différentiel, un plan d'investigation et un plan thérapeutique. Pour rendre la situation d'évaluation authentique, le patient est un patient véritable connu du professeur, le problème est un problème clinique véritable et courant que le résident est susceptible d'être appelé à résoudre dans une pratique professionnelle ultérieure.

La grille d'évaluation est une grille d'évaluation de type échelle descriptive globale (annexe V) avec neuf items décrivant les tâches professionnelles qui seront évaluées, et quatre niveaux de standards de performance (allant de « inadéquat » ou novice à « excellent » représentant la compétence professionnelle). Chaque standard de performance a des critères descriptifs décrivant ce qui est acceptable et ce qui n'est pas acceptable. Le contenu de cette grille a été pré-validé à une étape antérieure (Nguyen, 2006b).

L'accent de l'activité évaluative sera mis à la fois sur le processus qui permet au résident de résoudre le problème clinique et sur son argumentation scientifique.

La rétroaction à la fin de l'observation directe dure entre 15 à 20 minutes et est faite sous forme de « co-évaluation » (terme utilisé par Allal (1999) et Laurier (2005) pour désigner un processus d'évaluation impliquant à la fois le professeur et l'étudiant) : le résident doit s'auto-évaluer avec la même grille que celle du professeur. Ce dernier l'évalue aussi, mais de façon indépendante. Une comparaison des résultats qui se trouvent

sur les deux grilles permet par la suite de cibler la rétroaction sur les points où il y a divergence d'opinion.

Au cours de cette rétroaction, le professeur cherche à comprendre le cheminement cognitif du résident, la discussion interactive est favorisée et le rôle du professeur est de renforcer ce que le résident fait de façon adéquate et d'améliorer ce qui est perçu comme «non-atteint».

II.2 Recrutement des professeurs, des résidents et des patients

Les professeurs ont été invités verbalement à participer à ce projet de recherche. Ils ont été recrutés selon leur réputation de dévouement à l'enseignement, leur expérience antérieure avec l'observation directe, leur connaissance du programme de formation en médecine interne et leur expertise clinique. Toutes les spécialités médicales ont été approchées pour recruter des professeurs. Le nombre de professeurs est fixé arbitrairement de telle façon que chacun a deux séances d'observation directe par mois. Ceci tient compte de la lourdeur de la tâche qui est ajoutée à leur travail quotidien. Chaque professeur recruté a accepté de participer au projet tout au long de l'expérimentation et fait des ÉFPC pour des résidents qui lui ont été assignés aléatoirement. Ils ont accepté aussi de répondre à la fin de l'expérimentation à des questions lors d'une entrevue semi-structurée permettant de recueillir leur perception sur l'utilité de cette nouvelle méthode d'évaluation.

Les résidents ont été recrutés par lettre circulaire à tous les résidents du programme de médecine interne de l'Université de Montréal, de première année de résidence (R-I) jusqu'à la troisième année de résidence (R-III). Les résidents des quatrième et cinquième années ne furent pas inclus puisqu'ils étaient en cours de préparation de leur examen de certification. Cela permettait d'éviter des biais de confusion potentielle entre les attentes des ÉFPC et les attentes de l'examen de

certification. Chaque résident s'est engagé à avoir une ÉFPC par mois, ce nombre a été décidé arbitrairement étant donné qu'aucune donnée n'était disponible sur le nombre idéal de sessions qu'un résident devrait avoir par mois en évaluation formative pour vérifier sa progression.

Les patients ont été choisis par les professeurs participants. Trois critères ont été retenus : que le patient ait un problème médical stabilisé (pour éviter des décompensations de son état de santé au cours de l'ÉFPC), qu'il soit bien connu du médecin évaluateur ou tout au moins que son problème médical soit connu du professeur, et qu'il présente un problème médical courant qu'un médecin spécialiste est susceptible d'avoir à traiter.

II.3 Considérations éthiques

L'ordre de l'assignation des résidents aux professeurs a été fait par une méthode aléatoire. Tous les résidents et les professeurs ont été au préalable rencontrés individuellement pour s'assurer qu'il n'existait pas de conflit de personnalité latent ou manifeste.

La participation des professeurs et des résidents a été entièrement volontaire sans aucune rémunération financière. Ils ont signé une formule d'acceptation mentionnant qu'ils pouvaient se retirer n'importe quand sans aucune conséquence et toute analyse de donnée sera non-nominale.

Tous les patients ont été avisés que leur participation à l'activité était entièrement pour des fins académiques et qu'une stricte confidentialité des détails cliniques leur était assurée. Ils pouvaient, s'ils le désiraient, recevoir un montant d'argent symbolique pour couvrir leurs frais de déplacement.

II.4 Plan de recherche

Tous les mois, un tirage au sort permettait d'assigner un résident à un professeur. Le professeur recevait alors l'adresse de courriel de l'étudiant pour établir le contact.

Chaque professeur et chaque étudiant sont entrés en contact direct et se sont mis d'accord sur la date, l'heure et l'endroit où aurait lieu leur ÉFPC.

Pour simuler le contexte de travail réel des stages cliniques, les professeurs et les résidents ont reçu au préalable un document décrivant la méthode d'ÉFPC, les grilles d'évaluation utilisées au cours de l'ÉFPC. De surcroît, le professeur a reçu un document décrivant en détails la façon d'effectuer une rétroaction. Ce dernier document intègre les articles de Brinko (1993) et de Ende (1983), et décrit étape par étape une rétroaction (annexe I et II).

L'expérimentation a eu lieu de février 2005 à juillet 2005. En juillet, après le déroulement de toutes les séances d'ÉFPC, deux séries d'entrevues semi-structurées ont eu lieu. La première consistait en des rencontres individuelles avec des résidents et le questionnaire portait sur la perception qu'il avait de l'ÉFPC comme méthode d'évaluation formative. La deuxième série d'entrevues consistait à rencontrer des professeurs pour recueillir leur perception de l'utilité de la méthode d'ÉFPC en comparant avec d'autres méthodes d'évaluation.

II.5 Plan d'analyse des données

L'analyse des données est en premier lieu réalisée par une méthode d'analyse descriptive des fréquences des scores de chaque item qui permet de voir si certaines compétences étaient mieux développées que d'autres.

En deuxième lieu, une analyse de variance non paramétrique pour mesures répétées permet de vérifier le changement dans l'évolution des résidents.

En troisième lieu, une analyse qualitative des annotations écrites sur les grilles d'observation permet de vérifier la relation entre les résultats quantitatifs et ceux écrits par les professeurs.

Finalement, les résultats de l'entrevue de conclusion permettent de connaître les perceptions que les professeurs ont de la méthode des ÉFPC.

III. Analyse des données et résultats

III.1 Profil des participants et des sessions d' ÉFPC

Au total, cinq professeurs ont accepté l'invitation et quatre ont décliné l'offre. Les cinq ayant décliné ont donné le manque de temps comme raison principale. Le chercheur s'est ajouté comme sixième membre à l'équipe des professeurs pour maintenir le nombre de sessions d'observation directe à un niveau acceptable. Le tableau II montre le profil des professeurs participants. Il y a des spécialistes en cardiologie, en gastro-entérologie, en neurologie, en médecine interne et en rhumatologie. Cela permettait d'avoir une variété des domaines d'expertise clinique des professeurs. Leur expérience dans l'enseignement est aussi variable, allant de deux ans à vingt ans. Deux des professeurs ont été des directeurs de programme et quatre ont déjà participé à des examens nationaux de certification de leur spécialité. Ils ont tous effectué des observations directes dans différents contextes (ECOS des étudiants du niveau pré-gradué et des évaluations périodiques des résidents). De plus, ils sont tous impliqués dans des tâches de gestion ou de recherche. Une telle variabilité permet de s'assurer de la représentativité de la réalité de la formation clinique et d'éviter les biais potentiels d'homogénéité des résultats que la

présence des professeurs au profil trop homogène entraînerait. Chacun a effectué environ 11 sessions d'ÉFPC (allant de 10 à 14).

Le tableau III montre les caractéristiques des résidents. De façon générale, les treize résidents qui ont accepté l'offre de participer se distribuent de façon égale, selon le genre (sept hommes et six femmes) et le niveau de formation.

Le tableau IV montre les caractéristiques des sessions d'ÉFPC: type de problèmes cliniques, difficulté des cas et durée des sessions.

Au total, 68 sessions d'ÉFPC ont eu lieu. Il y en a eu 13 pour les périodes 1, 2, 3, 6; 11 pour la période 4; et 5 pour la période 5. La période 5 a été particulièrement difficile parce qu'il y a eu de multiples congés annuels et des absences pour des congrès à l'étranger. Le type de pathologie variait beaucoup, avec une prédominance de problèmes de médecine interne. Comme deux évaluateurs étaient des spécialistes en médecine interne, cela explique cette prédominance.

Les niveaux de difficulté estimés par tous les professeurs sont distribués en grande majorité comme «modérément difficile» à 59 %, avec 28 % de problèmes cliniques estimés facile et 12 % estimés difficile.

La moyenne de la durée totale des sessions a été de 36 minutes (avec un écart type de 13 minutes), la majorité des sessions pour lesquelles nous avons des données ont duré moins d'une heure (93 %), avec une durée de rétroaction d'en moyenne 29 minutes (écart type de 11 minutes) dépassant 15 minutes dans 91 % des sessions.

Quant à l'expérience clinique des résidents préalable à la session d'ÉFPC, à l'exception de la période 5, entre 77 % à 100 % des résidents avaient fait les stages pertinents avant la session. Ils avaient par ailleurs acquis un bagage minimum de connaissances de ces domaines lors de leur formation pré-graduée en médecine.

III.2 Consistance interne de la grille d' ÉFPC

L'étude de la consistance interne des items par l'analyse du coefficient alpha de Cronbach montre qu'il y a une amélioration au cours de l'utilisation de la grille d'ÉFPC par les professeurs et ce coefficient atteint le niveau de 0,81 au quatrième mois d'utilisation. Par contre, ces résultats doivent être interprétés avec prudence puisque le nombre d'items est petit, de même que le nombre d'étudiants. L'utilisation de la grille par des résidents montre aussi un coefficient alpha qui atteint 0,8 dès le premier mois et comme avec les professeurs, ce coefficient chute à 0,5 à la période 5 où la moitié des sessions n'ont pas eu lieu.

III.3 Étude de la progression des performances

Le tableau V rapporte l'analyse de variance non paramétrique pour mesures répétées reflétant les performances au début et à la fin. Cette analyse vérifie la progression de la compétence clinique des résidents. La moyenne des performances des sessions de la période 1 et 2 représente les scores de performance au début, cela permet de s'assurer que cette performance reflète le jugement de deux évaluateurs différents. Nous avons fait la moyenne des performances des sessions des périodes 4 et 6 pour établir les scores de performance de la fin et ont omis les scores de la période 5 parce qu'il n'y a pas eu assez de sessions. La compétence clinique globale est définie comme la somme des neuf sous-compétences, ceci pour souligner que toutes ces sous-compétences ont une importance égale.

L'analyse montre une différence statistiquement significative dans six des neuf sous-compétences : le questionnaire, le souci de vérifier l'impact psychosocial de la maladie, l'examen physique, la discussion avec le patient, les techniques d'entrevue et le plan d'investigation. D'autres sous-compétences, même si on note une tendance à leur amélioration, n'atteignent pas un niveau de différence statistiquement significatif. Cela

concerne la présentation de cas, le diagnostic différentiel et le plan de thérapie proposé. La différence est aussi statistiquement significative pour la compétence clinique globale.

L'étude Post Hoc Scheffé à comparaison multiple concernant la compétence clinique globale révèle une différence significative entre le niveau de résident I et résident III. Par contre, la différence entre le niveau de résident II avec celui de I ou de III n'est pas significative.

L'analyse descriptive de fréquence du tableau VI donne des détails sur le pourcentage de fréquence entre les niveaux « novice » (1 et 2) et « compétent » (3 et 4). Du questionnaire, on peut noter que des 46 % jugés comme ayant un niveau « novice » au début, seulement 25 % restent à ce niveau après six mois. Pour l'examen physique, nous retrouvons une amélioration de 25 % et ceux qui sont jugés comme n'ayant pas atteint le niveau de « compétent », se retrouvent au niveau 2 (« acceptable mais certaines améliorations sont nécessaires »).

Les sous-compétences telles que le questionnaire sur l'impact psychosocial et la discussion des informations cliniques avec le patient se sont aussi nettement améliorées : les 20 % et 39 % au début ayant un niveau « acceptable » ont progressé à 64 % et 88 %.

Par contre, pour les quatre dernières sous-compétences, les résidents ont un niveau de performance de base élevé de telle façon que la progression, même si elle est présente, est moins remarquable : tous ont un niveau de performance jugé acceptable au début (allant de 42 % à 68 %) et ont progressé jusqu'à 84 % (52 % - 84 %).

Il est notable que 80 % des performances concernant les techniques d'entrevue ont été jugées dès le début comme étant à un niveau acceptable.

Le profil de progression sur six mois de chaque sous-compétence pour les trois niveaux de résidence est reproduit sur les figures 1 à 9. La variabilité des performances

au cours de six mois nous permet de voir que les professeurs n'ont pas tendance à coter tous les résidents de la même façon au début et à la fin, ni d'une sous-compétence à une autre. Cette variabilité confirme de plus que les résidents n'évoluent pas de la même façon. La figure 10 montre par contre qu'en moyenne, la performance des résidents pour chaque sous-compétence progresse au cours de la période d'expérimentation. Et la figure 11 illustre la progression de la compétence clinique globale.

III.4 Analyse des annotations sur la grille d'ÉFPC

L'analyse des annotations sur la grille d'ÉFPC a été effectuée par la technique d'encodage avec le logiciel ATLAS.ti (V5.0). Des familles de codes représentant des annotations sur les domaines "à améliorer», «à maintenir» ont été créés. L'ensemble des suggestions pédagogiques sont regroupées et une famille de codes sur les cibles de la rétroaction a aussi été créée.

Le tableau VII reproduit ces résultats; les fréquences représentent le nombre de fois où ces informations ont été notées par rapport au total des commentaires.

Les annotations semblent refléter les soucis des professeurs, identifient les domaines où la performance de base des résidents ont été les plus faibles et portent en grande partie sur le questionnaire et l'examen physique. La majorité des commentaires qui ont trait aux techniques d'entrevue, au contact avec le patient et aux discussions d'informations cliniques notent les attitudes positives des résidents et les encouragent à les maintenir.

Les commentaires portent aussi sur la présentation de cas et le diagnostic différentiel, et les expressions fréquemment utilisées sont «synthétique» et «priorisation».

Finalement, les domaines où très peu de commentaires ont été faits sont ceux de l'investigation et du traitement.

Un total de 30 suggestions pédagogiques a été relevé. Elles sont très diversifiées, les principales portent sur des suggestions de lecture spécifique (de sémiologie ou de sujets cliniques), et sur des pratiques d'examen physique sous observation directe.

Les séances de rétroaction, telles que rapportées par les professeurs, portent sur la discussion et l'enseignement des techniques de questionnaire et d'examen physique. À quelques reprises, les professeurs soulignent l'importance des impacts psychosociaux de la maladie sur la vie du patient. Ces rétroactions mettent de plus l'accent sur l'intégration des informations pour mieux faire tel que l'illustrent les commentaires :

... «examen physique systématique et orienté»

...«examen neurologique plus orienté»

...«questionnaire avec rigueur ...»

... «diagnostic différentiel avec intégration des données ...»

III.5 Perception des professeurs de l'ÉFPC

L'entrevue avec les professeurs s'est déroulée en deux temps. La première rencontre était une entrevue de groupe avec quatre professeurs. Deux professeurs ont été retenus par des activités cliniques à la dernière minute et nous avons donc rencontré individuellement chacun d'eux en utilisant le même questionnaire .

Il s'agissait d'une entrevue semi-structurée où une liste de questions a été préparée préalablement pour obtenir la perception des professeurs sur l'utilité des ÉFPC sous l'angle de :

1. Ce que les ÉFPC apportent dans l'évaluation des étudiants

2. Le format de l'ÉFPC et les modifications recommandées pour la rendre acceptable

Les réponses recueillies lors de l'entrevue du groupe ont servi comme informations à être validées lors de l'entrevue individuelle avec les deux professeurs qui n'ont pu participer à la discussion.

La transcription de ces entrevues a été faite textuellement par une tierce personne et elle a été encodée en deux temps. Le premier encodage l'a été sous forme de codage très spécifique. Le deuxième sert à regrouper les thèmes qui se recoupent. Une analyse des segments non encodés sert à vérifier s'il n'y a pas eu de pertes d'information.

Cette analyse qualitative a donné lieu essentiellement à deux séries de commentaires : 18 négatifs et 86 positifs pour un total de 104 .

Les commentaires négatifs ont deux aspects :

1. Sur la méthode : La majorité des commentaires porte sur le «temps nécessaire pour faire l'évaluation»:
(11 commentaires)

« ... ça demande beaucoup de disponibilité ...

... si on veut bien faire, ... il faut y mettre beaucoup de temps. »

Le deuxième problème qui a été source d'irritation est «l'organisation difficile» :

« ... des fois, je fais venir un patient, et le résident ne vient pas parce qu'il est pris à l'hôpital, ... c'est difficile ... »

Des suggestions ont été proposées pour corriger ces problèmes :

« ...cette tâche doit être le travail de tout le service
... »

« ... ce travail doit être rémunéré ... »

2. Sur la grille : certains se plaignent de l'aspect restreignant :

(7 commentaires) « ... la grille porte à confusion ... »

« ... il fait certaines manœuvres et pas d'autres et ne
sais pas où cocher, alors j'ai écrit des
commentaires... ? »

et aussi :

« ...des fois, ça ne s'applique pas à certains cas.
J'aurais aimé mieux un peu de latitude ... »

Des commentaires positifs, trois aspects ont fait consensus parmi les professeurs :

1. Sur la rétroaction qu'ils ont donnée suite à l'évaluation de l'étudiant :

Soit parce qu'ils sont certains de ce qu'ils ont vu et qu'ils ont pu donner un
"feedback" adéquat :

« ...on était là pendant presque une heure ... on a tout vu ... toutes les
choses que j'avais à dire, je leur ai dites ... »

soit parce qu'ils ont été convaincus, que sur un plan individuel, ils ont
contribué à l'apprentissage de l'étudiant :

« ... quelques fois, je suis allée les rescaper et le "feedback" a été beaucoup plus facile ... »

« ...mes interventions ont été efficaces, que ce soit du côté positif ou négatif ... »

ou soit parce que le "feedback" a été reçu positivement :

« ... les résidents aussi étaient très ouverts, je ne les ai pas sentis sur la défensive ... »

« ... la rétroaction était quand même bien prise ... »

2. Sur la méthode, ils ont tous été étonnés, au cours de ces ÉFPC, de constater certaines forces et faiblesses des résidents :

« ... les premiers résidents que j'ai passés; j'étais quasiment étonnée. Je me dis : ouf, je ne vais pas être sévère ... »

« ... pas seulement aptitudes, mais les attitudes beaucoup ... »

« ... les détails, il y en a pour les fins, les fous ... à n'en plus finir ... »

3. Et concernant la grille, malgré des commentaires sur la confusion parfois retrouvés sur la grille, ils reconnaissent tous que la grille d'évaluation les a aidés à faciliter l'évaluation :

« ...on s'attarde pas juste sur ce qui est excellent, mais on explique aussi c'est quoi le "non-satisfaisant"... »

« ... ce qui m'a aidé, c'est quand je dois leur dire c'est bon et c'est pas bon ... et quand le résident conteste, j'ai lu les critères et j'ai dit, voilà ... »

De plus, même si tous les professeurs ont trouvé que les ÉFPC leur prennent beaucoup de temps, ils ont apprécié cette méthode puisqu'elle leur permet de :

« ... évaluer le résident en entier, et c'est vrai que cela permet de mettre le doigt dès le moment où cela commence à moins bien aller. »

« ... même détecter un résident odieux avec le patient, même si aucune évaluation n'en mentionne. »

« ... évaluer par une méthode pas très compliquée. »

« ... profiter pour montrer les choses ... »

« ... aller plus loin quand il y a un problème, si le résident est intéressé ... »

IV. Discussion et conclusion

L'analyse des résultats nous semble confirmer qu'une évaluation formative par l'approche des ÉFPC pour des résidents en médecine interne, des niveaux I à III pourrait les aider à progresser au cours des six mois de l'expérience.

Concernant l'aspect de contenu, comme aucun professeur ne critique les neuf sous-compétences, ni les quatre niveaux de standards de performance, cela nous rassure sur le contenu de la méthode d'EFPC pour évaluer les compétences cliniques.

Cette étude, en adoptant l'évaluation authentique comme approche, montre que la méthode des ÉFPC semble respecter la validité de structure, de substance et de conséquence, trois des six aspects de validité que Messick (1993; 1994) propose dans sa discussion sur la validité. La structure interne de l'outil de mesure, selon lui, vise surtout la comparabilité de performance entre les individus ou un groupe d'individus avec des performances désirées du domaine évalué. À cet effet, les standards de performance qui figurent sur la grille d'ÉFPC semblent alors permettre aux professeurs de comparer les performances de chaque étudiant aux différents niveaux de performances pré-établis. L'aspect de substance de la validité s'adresse quant à lui, aux besoins de s'assurer que les situations évaluées représentent bien les performances désirées. Messick insiste pour que les mesures dépassent une simple représentation de données des domaines mesurés. L'analyse qualitative des perceptions des professeurs démontre que ces derniers reconnaissent que les démarches cliniques dans lesquelles se sont engagés les étudiants représentent bien le travail clinique habituel. Quant à l'aspect de conséquence, Messick souligne l'importance de respecter, lors de l'interprétation des scores, l'intention avec laquelle des scores ont été attribués. Nos données démontrent alors que les professeurs ont profité des séances d'ÉFPC pour évaluer dans un but d'enseigner. Ils ont surtout mis l'accent sur aider l'étudiant à progresser.

L'aspect de la validité du construit ne se vérifie dans cette étude que par le fait que les ÉFPC permettent de distinguer les niveaux des résidents I et III quant à leur compétence clinique globale. Le trop petit nombre des résidents de chaque niveau ne permet pas d'établir une différence significative entre les R-II et les deux autres niveaux.

Parmi les neuf sous-compétences évaluées qui représentent le travail clinique habituel d'un médecin lors d'une première entrevue avec un patient, les six où les résidents ont le moins bien performé au début (le questionnaire du patient, la recherche des impacts psychosociaux de la maladie, l'examen physique, les discussions avec le patient sur leur problème, la technique d'entrevue et le diagnostic différentiel) sont aussi

celles où il y a eu une grande amélioration. Sans vouloir tirer une conclusion hâtive, il est fort possible que la rétroaction a joué un rôle déterminant. Ainsi, lorsque nous regardons la rétroaction, nous constatons que l'emphasis a été mise sur le questionnaire, l'examen physique, le contact avec le patient et le diagnostic différentiel. Cela explique fort probablement la progression significative des résidents. De la même façon, comme moins d'accent a été mis sur les domaines de l'investigation et du traitement, la progression des résidents constatée dans ces deux domaines est moins significative.

Cela confirme le rôle central de la rétroaction dans l'évaluation formative telle qu'a rapporté Hattie (1987) dans sa méta-analyse sur les facteurs qui influencent l'apprentissage et donne raison aux auteurs qui insistent pour dire que la clé à un apprentissage significatif passe par la rétroaction lors de l'évaluation formative des apprentissages (Hattie, 1998; Black, 1998 et Rushton, 2005).

Les données recueillies par les ÉFPC ont aussi une consistance interne acceptable dans la limite de cette recherche et assurent une certaine fidélité des mesures recueillies lors des sessions d'observation directe pour des performances complexes. Par contre, si cette méthode devait être utilisée à des fins sommatives, des recherches plus poussées de corrélation inter-juge pour les mêmes séances pourraient être nécessaires. Même si la grille a encore certaines lacunes telles que des confusions quant à la note à accorder dans certaines situations, les professeurs ont tous trouvé, après six mois d'utilisation, qu'elle ne nécessite que peu de modifications. De plus, dans l'esprit d'une évaluation formative, le consensus des professeurs est que cette grille fournit des points d'ancrage nécessaires pour lancer les discussions de rétroaction.

Le nombre limité de suggestions pédagogiques où les lectures spécifiques demeurent un des conseils principaux est par contre surprenant. Il est à se demander si pour les professeurs, des difficultés de performance ne découlent pas directement d'un manque de connaissances spécifiques aux problèmes présentés par le patient. D'autres

conseils plus cohérents consistent à inciter des résidents à avoir d'autres sessions d'ÉFPC, même si les professeurs ont été unanimes à trouver que le temps reste un facteur limitant important.

Finalement, un effet bénéfique inattendu des ÉFPC est que les professeurs, sans aucune formation préalable ni demandes spécifiques, ont spontanément profité de la durée des ÉFPC pour donner un enseignement individualisé et qu'ils ressentent une grande satisfaction d'avoir contribué à la formation des résidents.

Cette étude est la première étude, à notre connaissance, en médecine, où une cohorte d'étudiants a été suivie longitudinalement avec une méthode d'évaluation formative dans un contexte authentique pendant une courte période de temps et où nous avons pu observer une progression significative dans le développement des compétences.

Il serait, par la suite, intéressant de vérifier la stabilité de l'impact de cette formation en comparant, une ou deux années plus tard, les compétences des résidents qui ont participé à la présente étude avec d'autres résidents qui n'y ont pas participé.

TABLEAU II : PROFIL DES PROFESSEURS-ÉVALUATEURS.

ANNÉES D'ENSEIGNEMENT		NOMBRE
	< 5 ans	2
	5 – 10 ans	2
	> 10 ans	2
TÂCHES EXCLUANT L'ENSEIGNEMENT		
• Gestion pédagogique	Responsable hospitalier	4
•	Chef de service	1
• Recherche		1
• Clinique		
	Temps plein (≥ 40 h/sem.)	4
	Temps partiel (< 40 h/sem.)	2

TABEAU III : PROFIL DES RÉSIDENTS

Niveau de résidence	Nombre	Pourcentage(%)
1 ^{re} année	4	31 %
2 ^e année	5	38 %
3 ^e année	4	31 %

TABLEAU IV : CARACTÉRISTIQUES DES SESSIONS D'ÉFPC

TYPE DE PATHOLOGIE	NOMBRE N=68(*)		POURCENTAGE(***) (%)
Cardiologie	11		16
Gastro-entérologie	5		7
Médecine interne	32		47
Neurologie	10		15
Rhumatologie	10		15
DIFFICULTÉ DE CAS			
Facile	19		28
Modérée	40		59
Difficile	8		12
DURÉE TOTALE DES SESSIONS			
	n=57 (**)		
< 30 min.	8		14
30 – 60 min.	45		79
> 60 min.	4		7
De rétroaction (***)			
	n=67 (****)		
< 15 min.	6		9
15 – 30 min.	43		64
> 30 min.	18		27

(*) seules 5 sessions d'ÉFPC ont eu lieu au 5^e mois

(**) des données de 11 sessions sont manquantes

(****) les données de une session sont manquantes

(****) les pourcentages sont arrondis à l'unité

TABLEAU V : COMPARAISON DES COMPÉTENCES AU DÉBUT ET À LA FIN

	Début n=13 Moyenne (D.S.)		Fin n =13. Moyenne (D.S.)		F(p)
	Sous-compétence 1 :	Questionnaire	2.57 (± 0.49)	3.08 (± 0.45)	11.1 (0.006)
	Sous-compétence 2 :	Questionnaire sur l'impact psychosocial	2.26 (± 0.56)	2.77 (± 0.72)	7.09 (0.021)
	Sous-compétence 3 :	L'examen physique	2.11 (± 0.68)	2.69 (± 0.59)	8.76 (0.012)
	Sous-compétence 4 :	Discussion avec le patient	2.12 (± 0.68)	3.19 (± 0.33)	12 (0.000)
	Sous-compétence 5 :	Technique d'entrevue	3.08 (± 0.45)	3.54 (± 0.48)	5.33 (0.027)
	Sous-compétence 6 :	Présentation de cas	2.42 (± 0.44)	2.58 (± 0.44)	1. (0.337)
	Sous-compétence 7 :	Diagnostic différentiel	2.42 (± 0.53)	2.58 (± 0.64)	0.64 (0.436)
	Sous-compétence 8 :	Plan d'investigation	2.50 (± 0.64)	2.88 (± 0.30)	5.04 (0.044)
	Sous-compétence 9 :	Plan thérapeutique	2.63 (± 0.61)	2.79 (± 0.26)	0.88 (0.368)
Compétence clinique globale		21.8 (± 2.72)	25.9 (± 2.81)	27 (0.000)	

TABLEAU VI : COMPARAISON DE L'ÉVOLUTION DES SOUS-COMPÉTENCES

	Début			Fin		
	Période 1 %	Période 2 %	Moyenne	Période 4 %	Période 6 %	Moyenne
Sous-compétence 1 : (questionnaire)	Niveaux 1 & 2 53	38	46	41	8	25
	Niveaux 3 & 4 46	61	54	58	92	75
Sous-compétence 2 : (questionnaire de l'impact psychosocial)	Niveaux 1 & 2 84	77	81	25	46	36
	Niveaux 3 & 4 15	23	19	75	54	64
Sous-compétence 3 : (examen physique)	Niveaux 1 & 2 62	77	69	42	46	44
	Niveaux 3 & 4 38	23	31	58	54	56
Sous-compétence 4 : (discussion avec le patient)	Niveaux 1 & 2 44	77	61	25	0	12
	Niveaux 3 & 4 56	23	39	75	100	88
Sous-compétence 5 : (technique d'entrevue)	Niveaux 1 & 2 39	0	19	17	0	8
	Niveaux 3 & 4 62	100	81	83	100	92
Sous-compétence 6 : (présentation de cas)	Niveaux 1 & 2 54	46	50	33	46	40
	Niveaux 3 & 4 46	54	50	67	54	60
Sous-compétence 7 : (diagnostic différentiel)	Niveaux 1 & 2 69	39	54	58	38	48
	Niveaux 3 & 4 31	54	42	42	62	52
Sous-compétence 8 : (plan d'investigation)	Niveaux 1 & 2 42	33	38	25	8	16
	Niveaux 3 & 4 58	67	62	75	92	84
Sous-compétence 9 : (plan thérapeutique)	Niveaux 1 & 2 40	25	32	42	8	25
	Niveaux 3 & 4 60	75	68	58	92	75

TABLEAU VII : ANALYSE QUALITATIVE DES ANNOTATIONS

	Domaines	Nombre n (%)	Exemples
1.	À améliorer	97	
1.1	Questionnaire	22 (23)	... « questionnaire plus pertinent : rechercher les éléments-clés » ... « questionner de manière plus rigoureuse et précise » ...
1.2	Examen physique	32 (33)	... « étoffer et approfondir l'examen physique ... » ... « structure de l'examen physique à améliorer ... »
1.3	Contact avec le patient	1 (1)	... « utiliser des termes plus simples pour donner des explications »
1.4	Présentation de cas	18 (17)	... « être plus structuré dans la présentation des l'histoire ... » ... « synthèse à améliorer ... »
1.5	Diagnostic différentiel	19 (20)	... « trop détaillé. À prioriser le diagnostic et à mieux argumenter » « diagnostic différentiel à mieux exprimer » ...
1.6	Investigation	3 (3)	... « investigation selon diagnostic différentiel » ...
1.7	Traitement	2 (2)	... « impact du traitement sur le patient à vérifier » ...
2.	À maintenir	52	
2.1	Questionnaire	21 (40)	... « histoire logique »
2.2	Examen physique	12 (23)	... « excellent examen physique : structuré et complet »
2.3	Contact avec le patient	15 (29)	... « très bon contact avec le patient » ...

Figure 1 : Évolution du questionnaire du problème principal

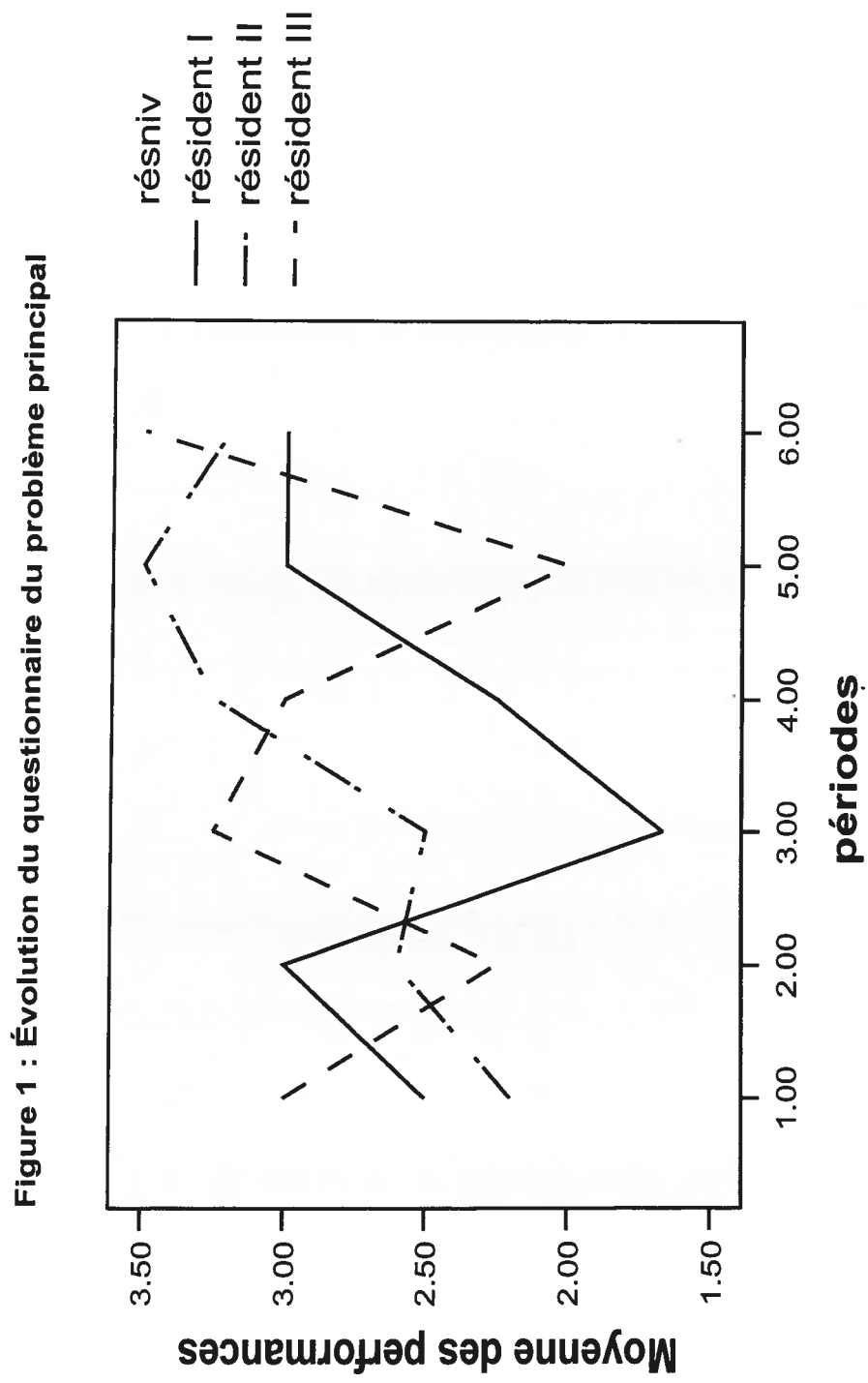


Figure 2 : Évolution du questionnaire sur l'impact psychosocial

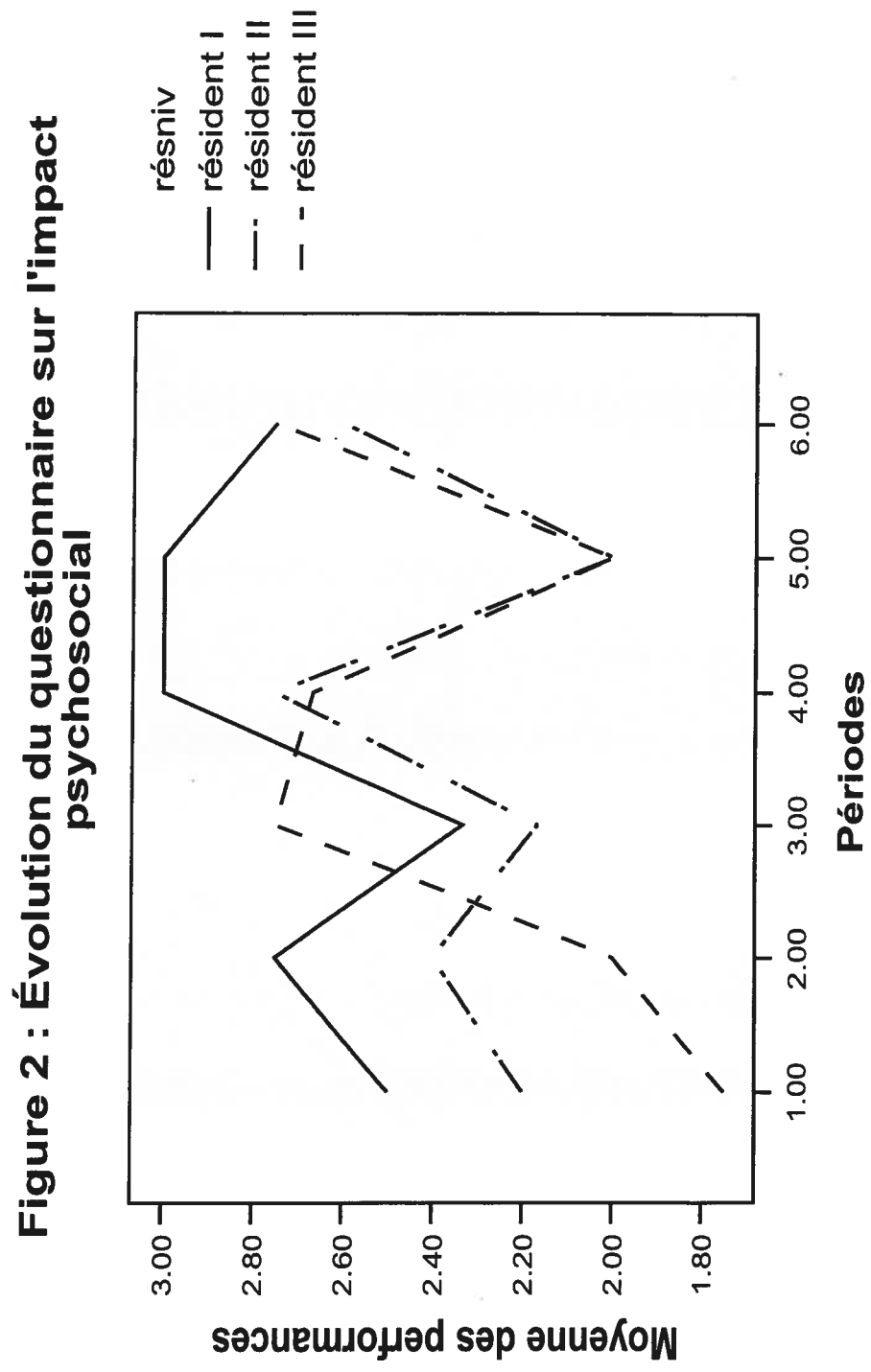


Figure 3 : Évolution de la technique de l'entrevue

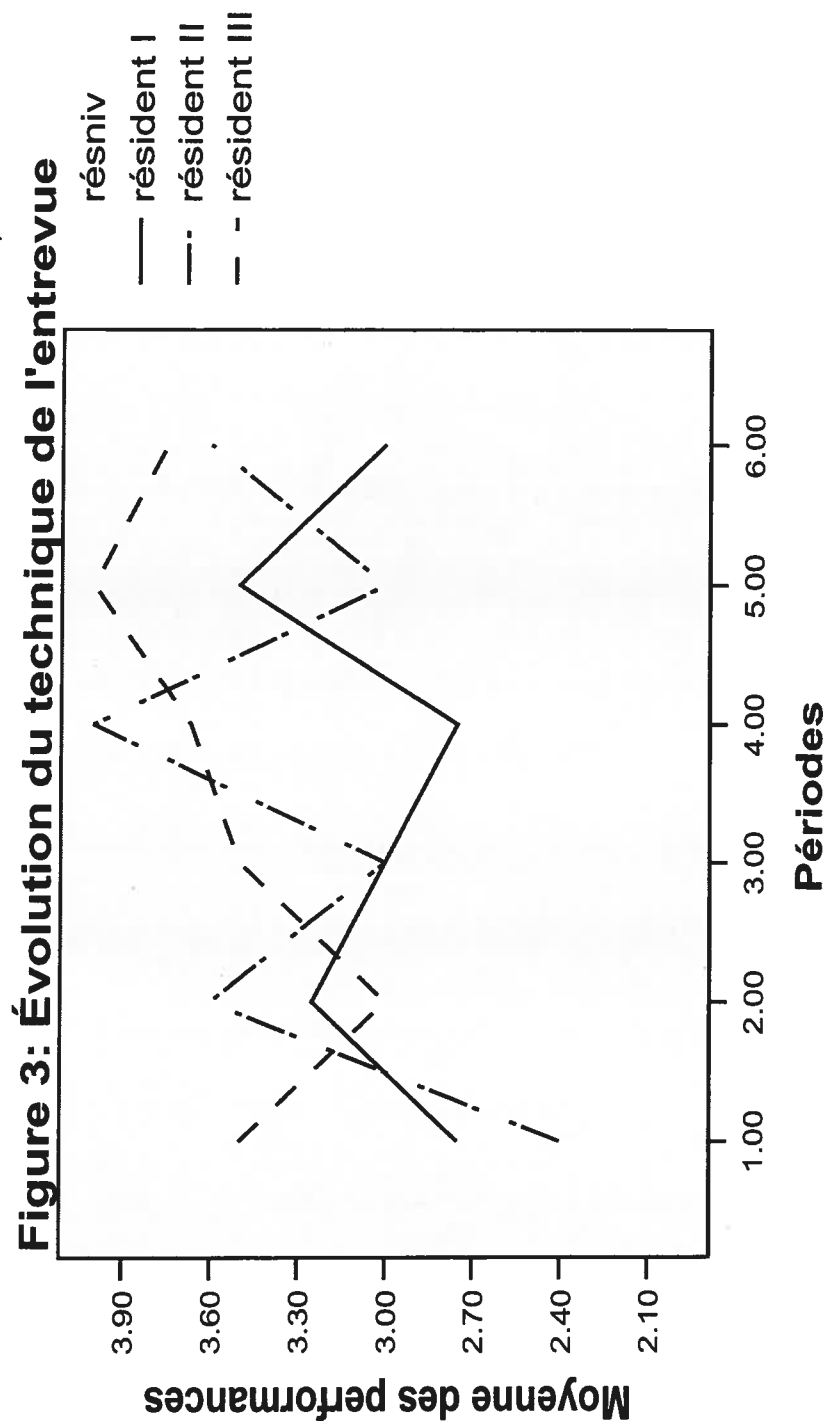


Figure 4 : Évolution de l'examen physique

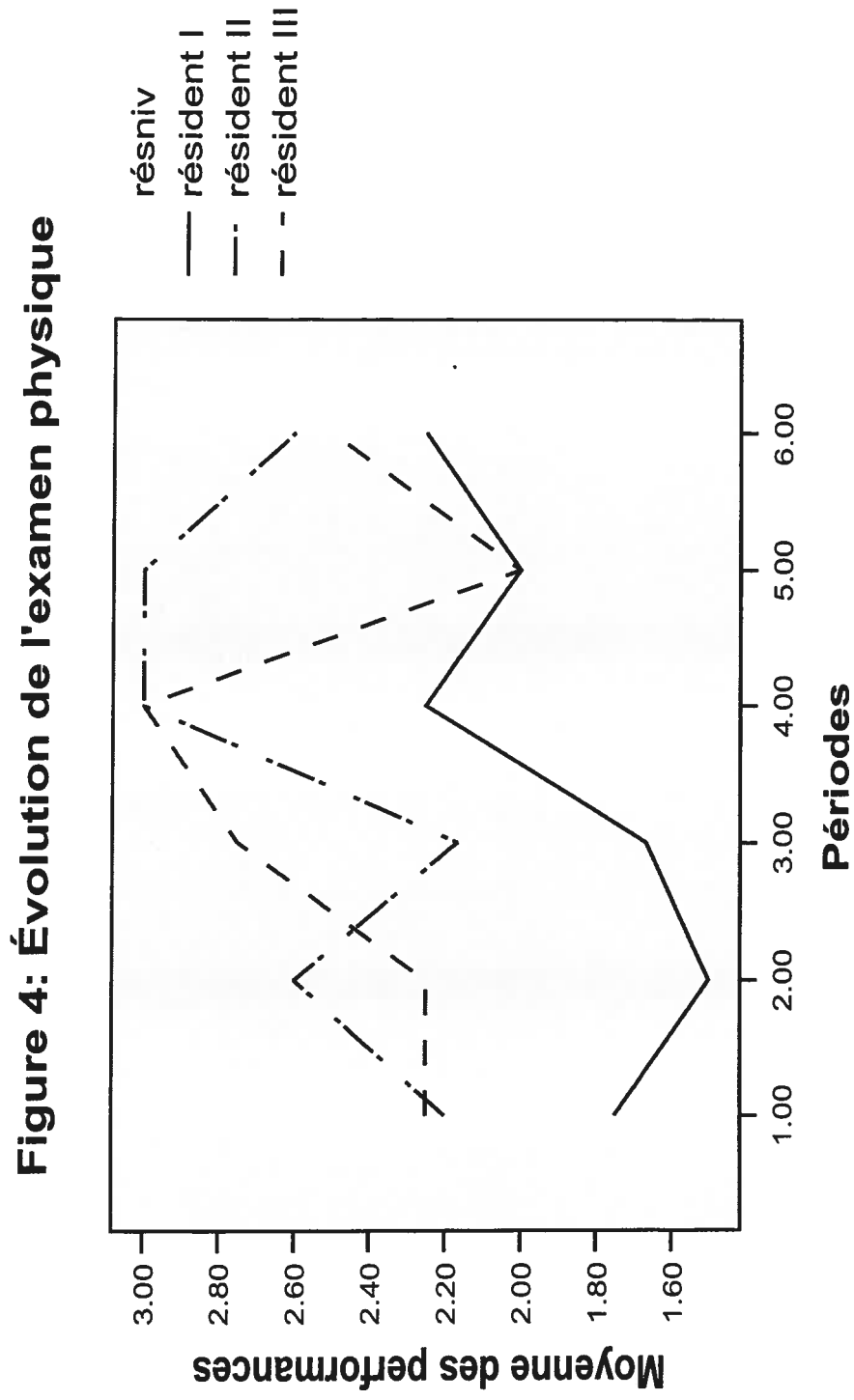


Figure 5 : Évolution de la discussion avec les patients

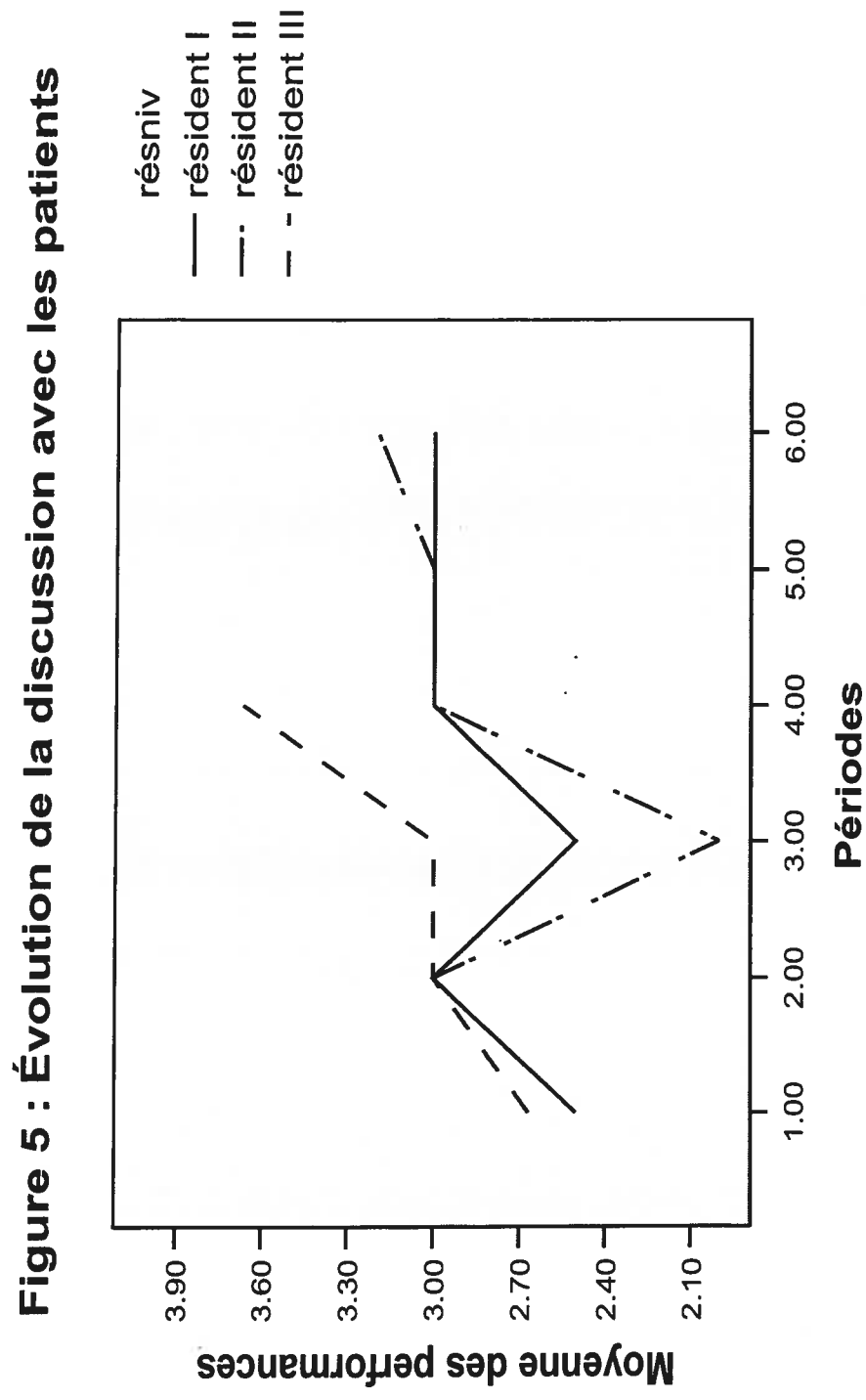


Figure 6 : Évolution de la présentation de cas

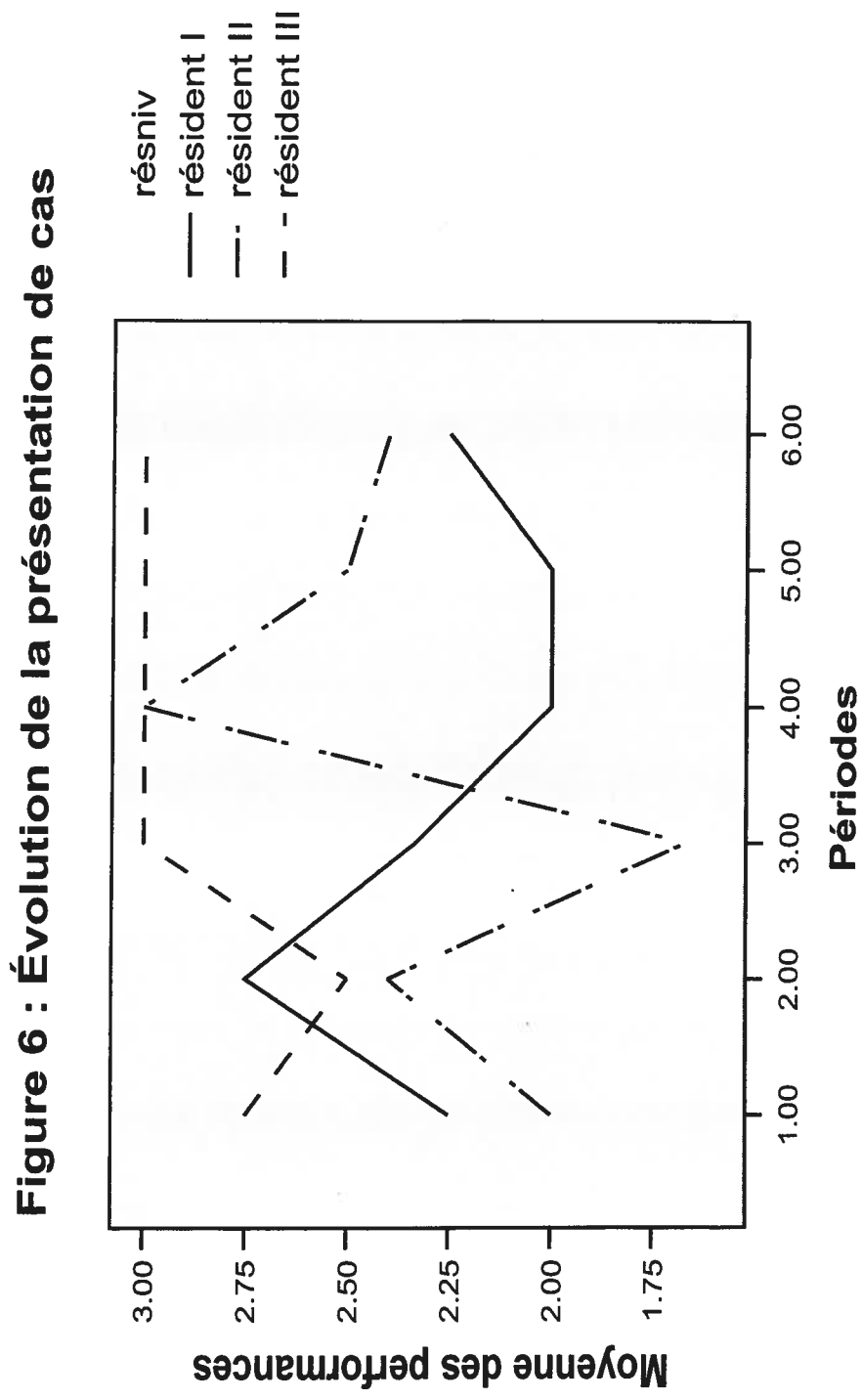


Figure 7 : Évolution du diagnostic différentiel

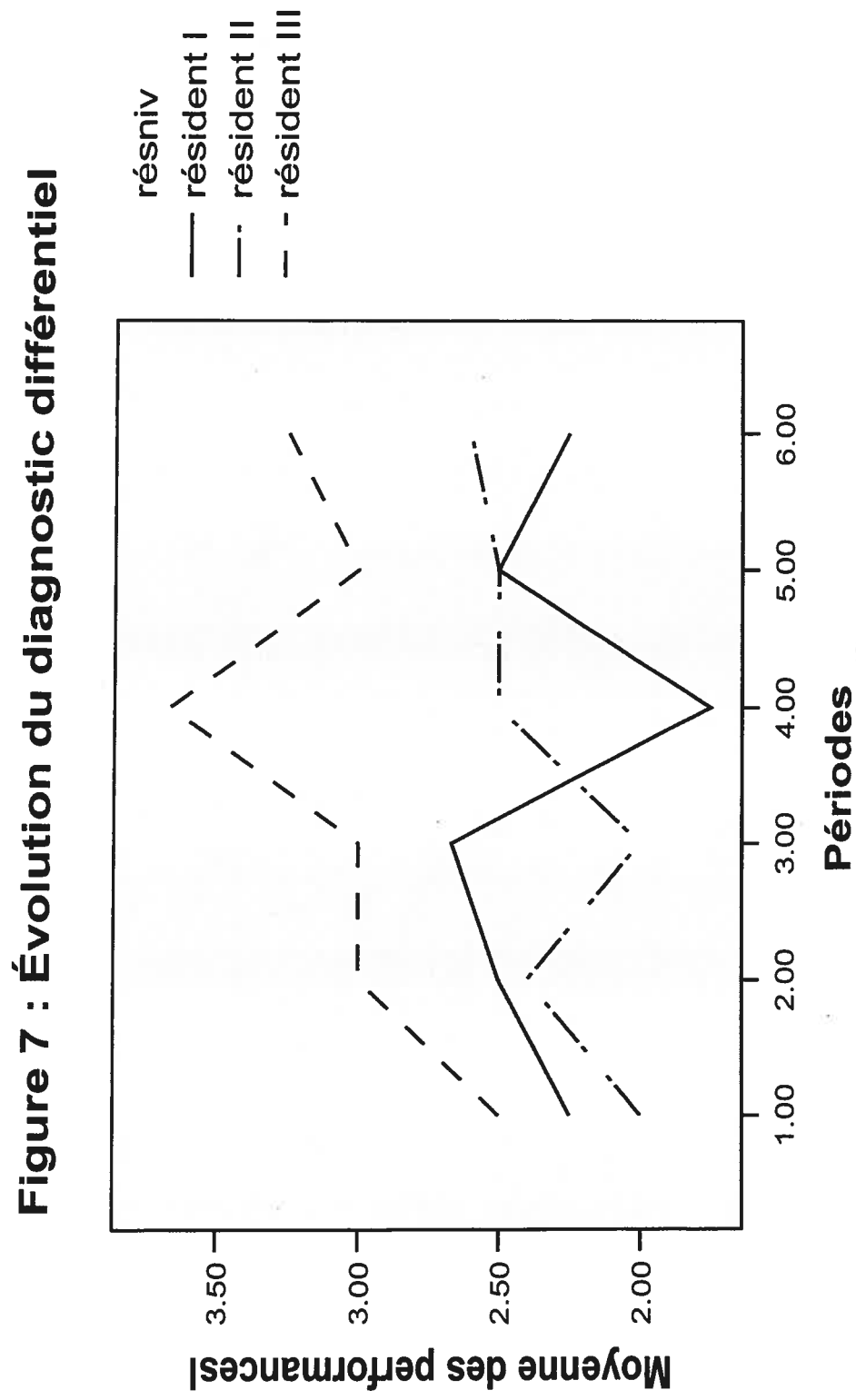


Figure 8 : Évolution du plan d'investigation

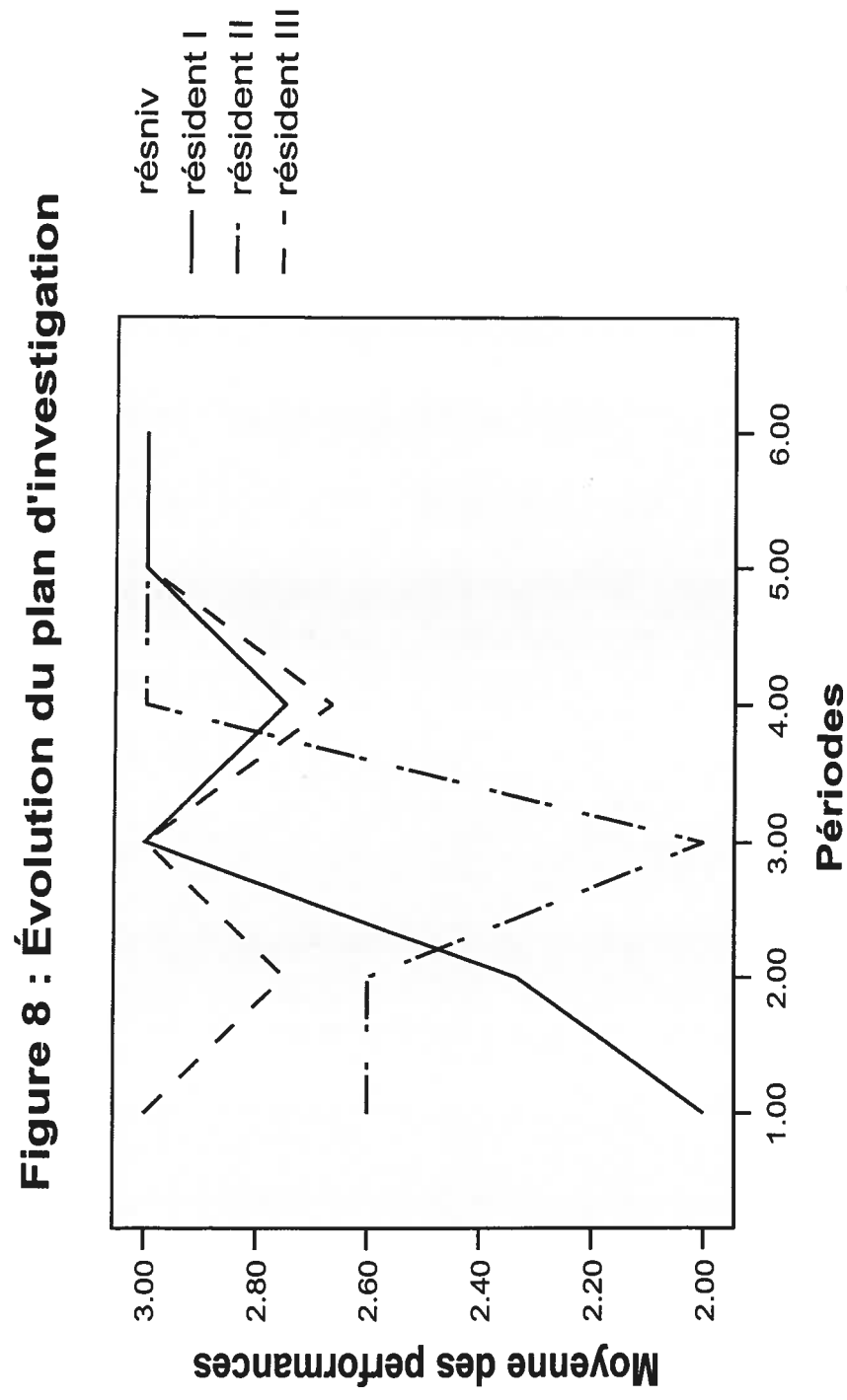


Figure 9 : Évolution du plan thérapeutique

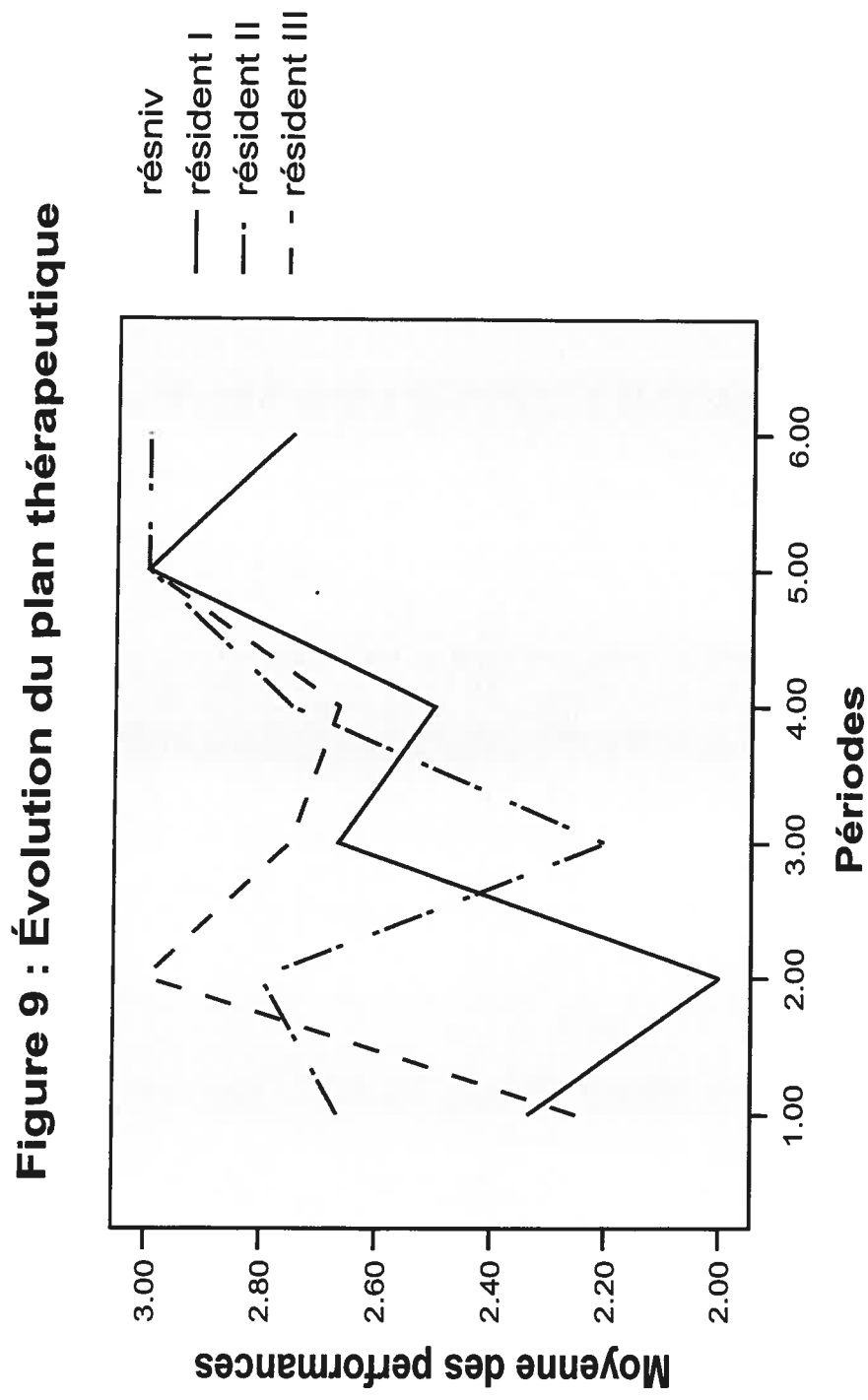


Figure 10 : compétences cliniques

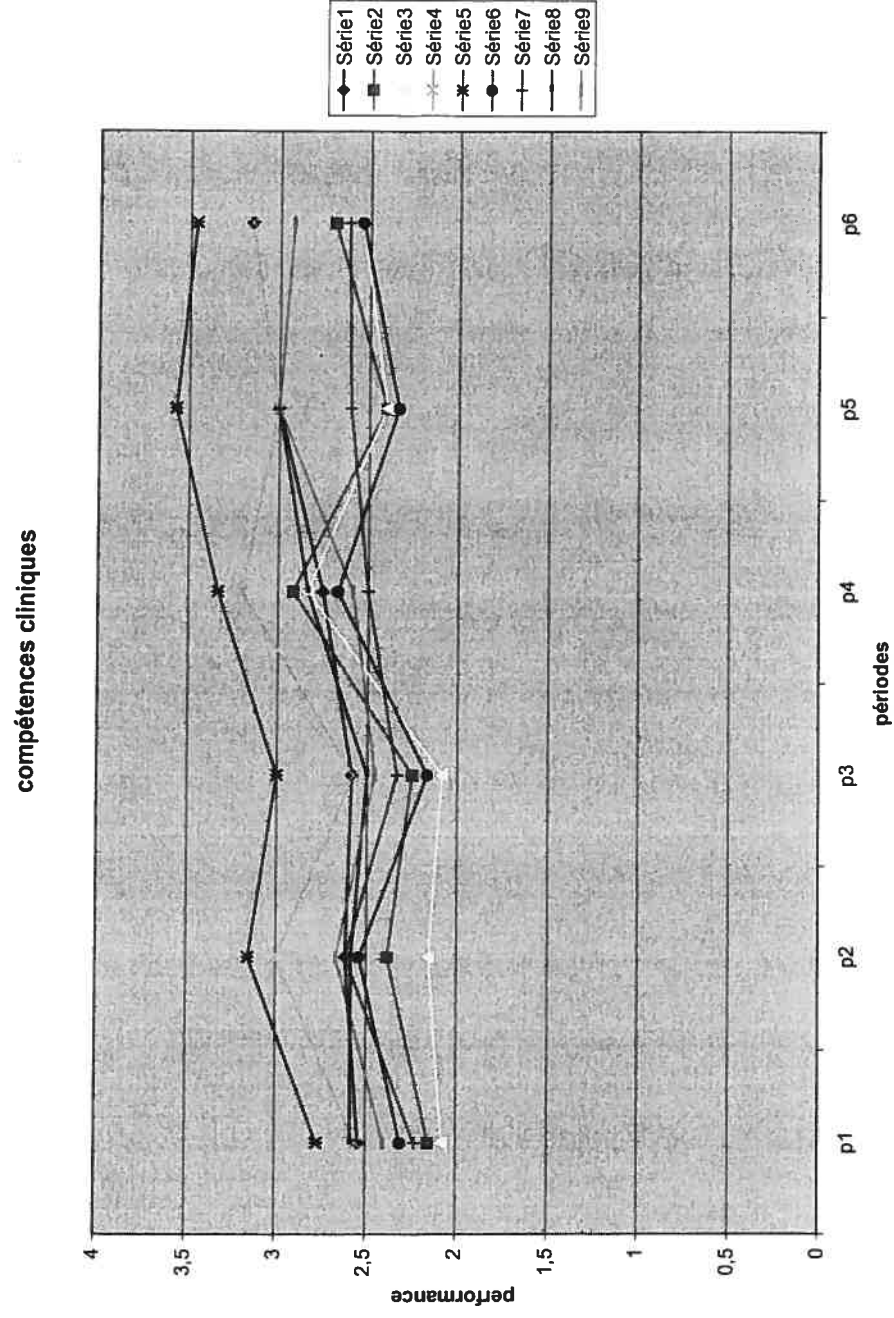
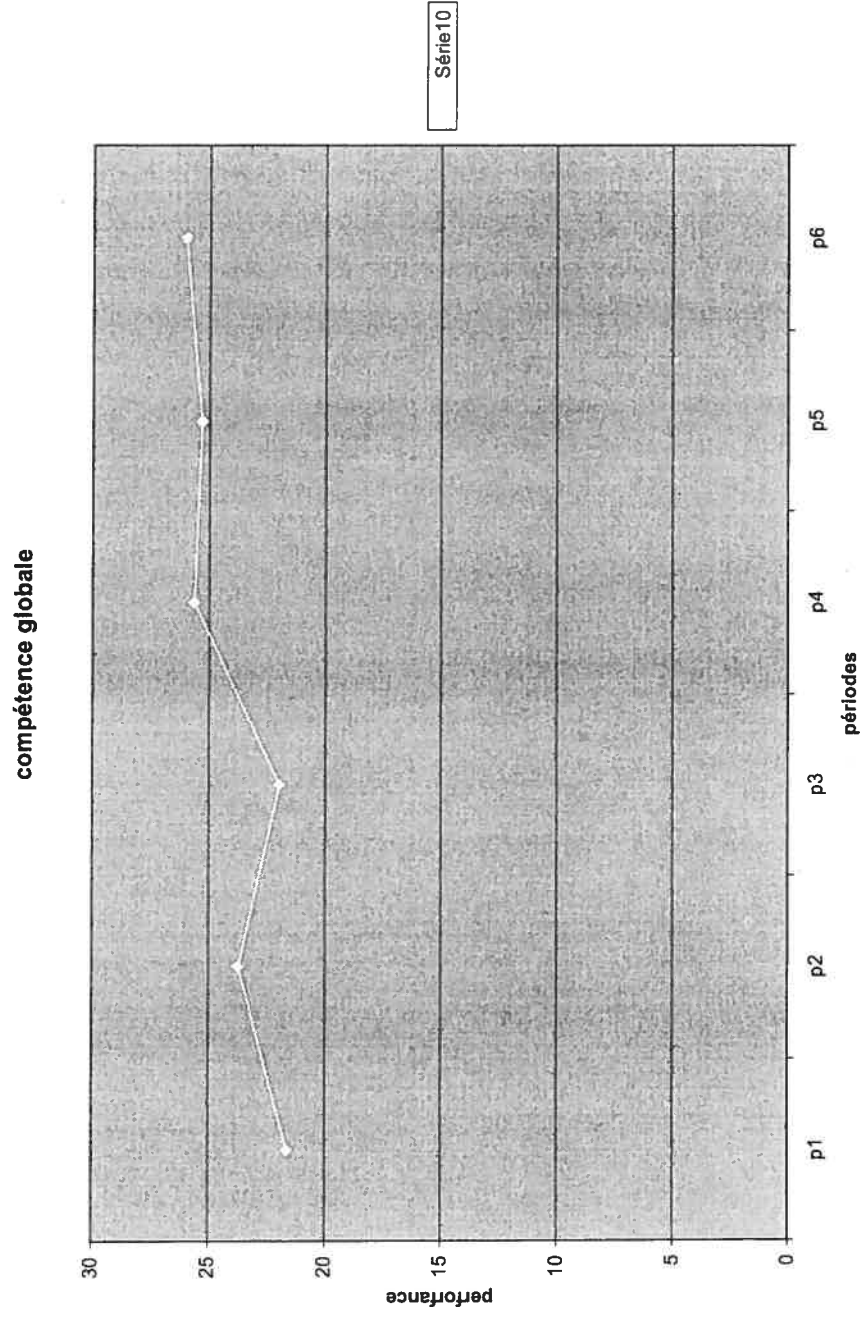


Figure 11 : compétence globale



RÉFÉRENCES :

Allal, L. (1999). Impliquer l'apprenant dans le processus d'évaluation; promesses et pièges de l'auto-évaluation. In C. Depover et B. Noël (dir.) : *L'évaluation des compétences et des processus cognitif*, pp 35-56. Bruxelles : De Boeck.

Battistone, M. J., Mine C., Sande M. A., Pangaro L. N., Hemmer P. A., Shomaker T. S. (2002). The feasibility and acceptability of implementing formal evaluation sessions and using descriptive vocabulary to assess student performance on a clinical clerkship. *Teaching & Learning in Medicine*, 14 (1) : 5-10.

Black, P., William D. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7-14.

Bloom, B. S., Hastings J. T., Madaus G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New-York : Mc Graw-Hill.

Brinko, K.T. (1993). The practice of giving feedback to improve teaching : what is effective ? *Journal of Higher Education*, 64 (5) : 574-593.

Day, S. C., Grosso L. J., Norcini J. J., Swanson D. B., Horne M. H. (1990). Residents' perception of evaluation procedures used by their training program. *Journal of General Internal Medicine*, 5 (5) : 421-426.

Ende, J. (1983). Feedback in clinical medical education. *JAMA*, 250 : 777-781.

Ferland, J. J. (1987). *Les grandes questions de la pédagogie médicale, perspective nord-américain*, chap. 1 à 5. Québec : Presses de l'Université Laval.

Frank, K. J. R., Jabbour M., Tugwell P. & al. (1996). Skills for the new millenium : report of the societal needs working group; CanMEDS 2000 Project. *Annals of Royal College of physicians and surgeons of Canada*, 29 : 206-216.

Gray, J. (1996). Global rating scales in residency education; *Academic Medicine*, 77 (Jan. Sup.), 5 : 55-63.

Hattie, J. (1987). Identifying the salient factors of a model of student learning : a synthesis of meta-analyses. *International journal of Educational Research*, 11 : 187-212.

Hattie, J. (1998). Assessment and classroom learning : a deductive approach. *Assessment in Education*, 5 (1) : 111-122.

Holm, H. A. (2002). Post-graduate Education. In G. R. Norman, C.P.M. Van der Vleuten, D. I. Newble (dir.), *International handbook of research in medical Education*, 381-413. United Kingdom : Kluwer Academic Publishers.

Holmboe, E. S., Hawkins R. E., Hammett T. W., Mackkrell-Gaglione M. (1996). Current methods in the evaluation of clinical competence. *Society of General Internal Medicine annual meeting* (1996).

Holmboe, E. S., Hawkin, R. E. (1998). Methods for evaluating the clinical competence of residents in Internal medicine : a review. *Annals of Internal Medicine*, 129- : 42-48.

Holmboe, E. S. (2004). Faculty and the observation of trainees' clinical skills : problems and opportunités. *Academic Medicine*, 79 (1) : 16-22.

Jolly, B. C. (2002). Faculty development for curricular implementation. In Norman G. R., Van der Vleuten C. P. M. & Newble D. I. (dir.), *International handbook of research in Medical Educatio*, 945-968. United Kingdom : Kluwer Academic Publishers.

Messick, S. (1993b). Validity. In R. L. Linn, *Educational measurement*, (3^e ed.), 13-104. New-York; American council on Education & Oryx Press.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2) : 13-23.

Nguyen, D.Q. (2006a). Approche par objectifs ou approche par compétences : quelles implications pour l'enseignement et pour l'évaluation dans la formation clinique ? Article qui sera soumis à la revue *Pédagogie médicale*.

Nguyen, D. Q. (2006b). Modélisation d'une grille d'évaluation formative des performances dans une perspective d'approche par compétences : un exemple de formation post-doctorale en Médecine Interne. Article qui sera soumis à la revue *Mesure & évaluation en éducation*.

Norcini, J. J., Blank L. L., Aznold J. K., Kimball H. R. (1995). The mini-CEX (Clinical Evaluation Exercise) : a preliminary investigation. *Annals of Internal Medicine*, 123 : 795-799.

Norcini, J. J., Blank L. L., Duffy S., Fortna G. (2003) : The mini-CEX : A method for assessing clinical skills. *Annals of Internal Medicine*, 138 : 476-481.

Pangaro, L. (1999). A new vocabulary and other innovations for improving descriptive in-training evaluations. *Academic Medicine*, 74 (11) : 1203-1207.

Rushton, A. (2005). Formative assessment; a key to deep learning. *Medical Teacher*, 27 (6) : 509-513.

Turnbull, J. , Gray J., MacFadyen J. (1998). Improving in-training evaluation programs. *J. Gen. Intern. Med.*, 13 : 316-328

Turnbull, J., Gray J., MacFadyen J., Van Barneveld C., Norman G. (2000). Clinical work sampling : a new approach to the problem of in-training evaluation. *Journal of General Internal Medicine*, 15 : 556-561.

Voorhees, R. A. (2001) Competency based learning models : a necessary future. In R.A.Voorhees (dir.), *Measuring what matters: competency-based learning models in higher education. New directions for institutional research*, (110) : 5-13. San-Francisco : Jossey-Bass.

Wakefield, J. (1985). Direct observation. In V. R. Newfield & G.R. Norman (dir.), *Assessing clinical competenc*. New-York : Springer Series on Medical Education, vol. 7, 51-70.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46 (7) : 41-47.

Wiggins, G. (1998). *Educative assessment : designing assessments to inform and improve student performance*. San-Francisco : Jossey-Bass.

Yorke, M. (1998). The management of assessment in Higher Education. *Assessment & Evaluation in Higher Education* : 23 (2) : 101-116.

Article 4

**Développement de la capacité d'auto-évaluation des
étudiants par une méthode d'évaluation authentique :
Résultats avec des ÉFPC**

Diem-Quyen Nguyen

Département d'administration et fondements de l'éducation
Faculté des sciences de l'éducation

Université de Montréal

© 2006

RÉSUMÉ

L'auto-évaluation est considérée comme une étape-clé dans le développement des compétences. Pour tous, c'est l'habileté qui favorise la formation tout au long de la vie professionnelle et qui permet de maintenir les compétences. Par contre, les études sur des stratégies pour développer l'auto-évaluation ont donné des résultats contradictoires.

Notre étude rapporte les résultats des Évaluations Formatives des Performances Cliniques. Il s'agit d'une expérience où les résidents se sont évalués avec des échelles descriptives globales où les standards de performance ont été explicités. Ils comparent de plus leur auto-évaluation à l'évaluation des professeurs pendant une période de six mois. Suite à cela, les résidents ont une rétroaction sous forme de discussion interactive avec leurs professeurs sur des points de divergence d'opinion.

Les résultats montrent que les résidents ont tendance à se surévaluer au début. De plus, à l'exception de l'examen physique, aucune corrélation significative n'est notable avec l'évaluation des professeurs. Par contre, à la fin, la corrélation entre les scores résultant de l'auto-évaluation et de l'évaluation par les professeurs montre une corrélation à 0,9 en ce qui concerne la compétence clinique globale. Des neuf sous-compétences composant cette compétence clinique globale, quatre ont atteint un niveau de corrélation qui dépasse 0,6 et trois sont situées au-dessus de 0,3. Seuls deux sous-compétences n'ont pas montré d'amélioration.

L'analyse qualitative montre que lorsqu'il y a un chevauchement entre les domaines de préoccupation des professeurs et des résidents avec une rétroaction ciblée, la capacité des résidents à s'auto-évaluer s'améliore. Nous discuterons enfin des implications pédagogiques de ces résultats.

Mots-clés

Auto-évaluation – Co-évaluation – Évaluation des compétences – Évaluation formative – Échelle descriptive globale –

ABSTRACT

Self-assessment is considered to be a key component in the development of clinical competence. For everyone, it underlines the lifelong learning process, which is an important component of professionalism. But research on the direction of strategies to develop self-assessment skills gives conflicting results.

In this study 13 residents were followed during six months. They have one session of Formative Clinical Performance Assessment per month and self-assessed themselves using a global descriptive scale, where standards of performance were clearly described. They compared their self-assessment results with their teacher's assessments and the resultant feedback was to be focussed on the areas of disagreement.

Our results show that, at the beginning, the residents overestimated their performances. And except for the physical examination sub-competence, there is no correlation between self-assessment scores and the teachers' scores. But, at the end of the experience, a level of correlation up to 0.9 was established for the global clinical competence. Four of the nine sub-competencies of this global clinical competence reach the higher level than 0.5 and three either maintain the 0.3 level or progress from a negative correlation to the positive level of 0.3. Only two sub-competencies show a decrease in the correlation. Qualitative analysis of all comments on the scales shows that where there is an overlap in the areas of concerns, either from residents, from teacher or from both of them, progress occurred in the self-assessment skills. Educational impacts of these results are discussed within the context of formative assessment.

Key words

Self assessment – Co-assessment – Competency assessment –
Formative assessment – Global descriptive scale.

I. Introduction

La formation médicale nord-américaine a résolument adopté l'approche par compétences depuis la fin des années 1990, avec l'obligation progressive faites aux facultés et écoles de médecine américaines d'incorporer les six compétences de "ACGME Outcome Project" (ACGME, 1999) dans leur programme de formation postdoctorale. Au Canada, les sept compétences des médecins spécialistes de "CANMEDS 2000" (Frank, 1996) font aussi partie des programmes de formation médicale postdoctorale d'études spécialisées. L'approche par compétences dans ces deux pays souligne l'importance d'intégrer des connaissances, des habiletés psychomotrices et affectives pour résoudre efficacement des problèmes cliniques. Les compétences telles que le professionnalisme, la communication, le maintien de compétence, la collaboration, la gestion de ressources et la promotion de la santé occupent de plus en plus une place importante dans la formation des futurs médecins.

Pour développer des compétences, quelles qu'elles soient, des auteurs tels que Scallion (2004), Wiggins (1998), Gagné (2005) et Huba et Freed (2000) suggèrent l'approche de l'évaluation authentique pour vérifier si l'étudiant est capable de résoudre des problèmes mal définis dans un contexte proche de son futur contexte de travail. Selon cette approche, des problèmes complexes, fictifs ou réels, sont soumis à l'étudiant qui, pour les résoudre, devrait proposer des solutions novatrices où la simple récitation des connaissances ne pourrait suffire. De plus, la résolution de ces problèmes fait appel à l'intégration des connaissances et des habiletés psychomotrices et affectives. L'évaluation, toujours selon l'approche de l'évaluation authentique, requiert une appréciation de la qualité du résultat et du processus de résolution du problème en prévoyant aussi un mécanisme de rétroaction.

Dans le contexte de l'évaluation formative des compétences, l'appréciation des résultats et du processus de résolution des problèmes par des jugements professionnels sert de base à la rétroaction; cette dernière faisant partie intégrale du processus. Le développement de l'auto-évaluation et de la métacognition semble de plus tout aussi essentiel selon Saddler (1998), Scallon (2004) et Allal (1999). Saddler (1998) écrivait : «... le but ultime de l'éducation n'est pas seulement de favoriser l'acquisition des connaissances mais aussi de rendre l'étudiant progressivement indépendant du professeur et être capable d'apprendre seul tout au long de sa vie ... ». Selon Allal (1999), Gipps (2003) et Scallon (2004) l'auto-évaluation est une étape vers le développement de la métacognition. Ces auteurs considèrent que l'auto-évaluation est l'évaluation de l'étudiant de ses propres performances en se servant d'un référentiel externe; tandis que la métacognition est la capacité de l'étudiant de connaître et de juger son propre processus d'apprentissage pour l'améliorer. D'autres auteurs avancent que le développement de la capacité à s'auto-évaluer est essentiel à un professionnel compétent dans son maintien de compétence¹ tout au long de sa carrière professionnelle (Moore, 1992; Gordon, 1992; Ward, 2002).

Mais, les études sur l'auto-évaluation rapportent des résultats parfois contradictoires. La plupart des auteurs cherchent à connaître la capacité à s'auto-évaluer en faisant des études corrélatives soit avec des évaluations des professeurs, soit avec celles des autres étudiants, et même parfois avec celles des patients simulés. La méta-analyse de Falchikov & Boud (1989) rapporte quelques-unes de ces études. Selon leurs données, il y a une grande variabilité des résultats de corrélation entre les scores de l'auto-évaluation et ceux des professeurs allant de -0,05 à 0,82. Cette première étude méta-analytique sur des études au sujet de l'auto-évaluation a mis en lumière une grande

¹ Peu de termes français traduisent bien l'esprit de l'expression "lifelong learning"; et les termes « maintien de compétence » ainsi que « formation tout au long de sa vie » seront utilisés pour représenter cette expression.

variabilité de calculs statistiques, de méthodologies de recherche, d'instruments de mesure et de la nature même de l'activité d'auto-évaluation. Premièrement, à propos des méthodes de calculs statistiques, il y en a trois principales. Certains résultats sont rapportés sous forme d'effets de taille ("effect-size calculation") qui sont des données statistiques descriptives; d'autres sont rapportées sous formes de proportion et donnent le degré d'accord «inter-juge» entre l'auto-évaluation et l'évaluation des professeurs. Finalement, la majorité des études rapportent les résultats sous forme de coefficient de corrélation, ce qui permet de mettre en rang les scores résultants de l'auto-évaluation et de l'évaluation des professeurs.

Parmi ces trois méthodes de calculs, les auteurs donnent leur préférence au coefficient de corrélation puisque dans le premier cas, aucun niveau de signifiante statistique n'est disponible; et dans le deuxième, on est obligé d'utiliser les notes du professeur comme un standard et donc la proportion de « succès » telle que jugée par le professeur a une valeur absolue de 100 %. Ce processus est du moins contestable puisque la corrélation inter-juge en soi n'atteint rarement le niveau de 1. Il faut cependant distinguer l'accord de la corrélation afin de mieux comprendre les résultats rapportés

Deuxièmement, au sujet des méthodologies de recherche, Falchikov et Boud (1989) s'interrogent sur le déroulement du processus d'auto-évaluation, où dans certaines études, les étudiants ont des notes additionnelle «...s'ils font un effort d'auto-évaluation ». D'autres appellent l'auto-évaluation, la note que l'étudiant s'accorde pour prédire ses notes réelles obtenues à la fin d'un cours (Herbert,1990). Dans l'ensemble, le terme « auto-évaluation » est utilisé dès que l'étudiant doit apprécier, soit son travail, soit le processus d'apprentissage qui lui permet d'atteindre le résultat final (Miller, 2003). Les termes «auto-évaluation» et « auto-notation » sont alors utilisés de façon interchangeable et ajoutent à la confusion.

Quant à l'instrument de mesure, il va de la liste de vérification avec des catégories binaires à l'échelle ordinale avec plusieurs catégories descriptives (Ward, 2002). Ward a aussi rapporté dans sa méta-analyse une grande variabilité des résultats et propose l'hypothèse selon laquelle, cette variabilité découle de l'absence d'un standard avec lequel on peut établir une corrélation.

Même avant cette conclusion de Ward, il y a eu deux autres études de Gordon en 1991 et 1992 pour tenter de comprendre les difficultés de l'auto-évaluation. Dans un premier article, Gordon a effectué une méta-analyse sur l'auto-évaluation en se limitant à la formation professionnelle des sciences de la santé. Il a alors tenté de vérifier la validité de l'auto-évaluation ainsi que la précision des mesures de l'auto-évaluation. Sa définition de la validité de l'auto-évaluation fait allusion aux jugements en se basant sur des critères appropriés tandis que la précision des mesures est basée sur une comparaison entre les scores découlant de l'auto-évaluation et ceux obtenus avec d'autres instruments de mesure validés. Ses analyses qui portent sur l'auto-évaluation des connaissances factuelles, des performances enregistrées en audio-visuel et celles obtenues aux évaluations de stage ont montré des niveaux de corrélation variables et inconstants. Les étudiants ont de plus une tendance à s'évaluer « globalement » avec peu de changement dans le temps. Dans le deuxième article, Gordon (1992) a tenté de comprendre les raisons de cette absence d'amélioration dans la capacité d'auto-évaluation en révisant des articles qui rapportent des dimensions qualitatives dans l'auto-évaluation. Il a ainsi démontré qu'il y a une certaine incohérence théorique concernant l'auto-évaluation à la lecture de certaines études. Quelques-unes de ces études utilisent l'auto-évaluation pour vérifier la perception de l'étudiant sur lui-même; d'autres pour vérifier l'auto-régulation et finalement, certaines autres l'utilisent pour vérifier l'efficacité de l'enseignement. Gordon propose alors de recentrer l'auto-évaluation sur les besoins de l'étudiant d'améliorer leur propre capacité à évaluer leur connaissance et leur performance. Il y

mentionne aussi que la clarification des standards d'évaluation devrait occuper une place importante dans l'auto-évaluation.

Ces confusions du but de l'auto-évaluation, du rôle de l'étudiant et des moyens mis en place pour favoriser le développement de l'auto-évaluation se retrouvent aussi dans un article de réflexion sur l'auto-évaluation de Allal (1999). Selon cet auteur, lorsque l'on veut impliquer l'étudiant dans sa propre évaluation, il existe un réel danger de confondre les rôles, les moyens et les buts. De plus, les complexités du contexte de résolution de problème ajoutent d'autres dimensions de difficultés à l'auto-évaluation.

Dans un effort pour améliorer la capacité d'auto-évaluation, Ward (2003) rapporte une expérience où les résidents en chirurgie évaluent leur propre technique opératoire après avoir visionné des enregistrements audio-visuels sur des exemples de performances désirables et indésirables et ceux enregistrant leur propre opération. Aucune amélioration de la capacité d'auto-évaluation n'est notable après le visionnement des enregistrements illustrant des standards de performance. Par contre, les résidents montrent une amélioration après avoir regardé l'enregistrement de leurs propres performances, avec une corrélation qui a passé de 0.5 à 0.63. Cependant, Blackstein (2004) n'a pas été capable de démontrer un impact positif sur la performance chirurgicale des résidents après qu'ils aient eu une rétroaction sous forme de visionnement des enregistrements audio-visuels de leur propre technique opératoire.

D'autres stratégies rapportées pour améliorer la capacité d'auto-évaluation ont aussi donné des résultats contradictoires, qu'il s'agisse de faire participer l'étudiant au processus d'établissement des standards d'évaluation (Orsmond, 2002), ou de former à l'évaluation avec des ateliers de discussion (Fallows, 2001) ou de rendre les grilles d'appréciation plus spécifiques (Miller, 2003). Les corrélations entre l'auto-évaluation et l'évaluation par les pairs sont tout aussi faibles. Seul une partie des données de Orsmond (2002) semble montrer une meilleure corrélation entre l'auto-évaluation et l'évaluation

des professeurs lorsque des exemples de standards de performances sont fournis aux étudiant avec une rétroaction interactive. Ils ont atteint un niveau de concordance de 50% à 80% comparativement à 34% à 47% lorsqu'il n'y a eu que de rétroaction entre les pairs. Ces résultats sèment le doute sur les stratégies pour développer l'auto-évaluation.

Le problème ne semble pas être isolé à l'activité de « voir soi-même performer » pour mieux s'évaluer ni à celle de compréhension simple des attentes. Est-ce dû à des problèmes intrinsèques de l'outil de mesure, notamment les échelles d'appréciation? Klenowski (1995) rapporte dans son étude exploratoire sur le processus d'auto-évaluation, des constatations selon lesquelles ce processus s'améliore quand les critères et standards sont négociés et discutés entre l'étudiant et le professeur et qu'il y a une interaction avec le professeur. Boud (2000) a aussi analysé le rôle de l'évaluation dans le développement de la capacité à l'apprentissage tout au long d'une vie et il va dans la même direction que Klenowski: une évaluation basée sur les standards est préférable à une évaluation normative puisque l'étudiant sera mieux informé quant à ce que l'on attend de lui. De savoir qu'il performe mieux que d'autres étudiants l'informe très peu sur ses véritables capacités. Est-ce possible alors que les recherches en développement de l'auto-évaluation rapportent des résultats divergents parce que, fondamentalement, l'étudiant a dû s'évaluer sans qu'il y ait eu de clarification suffisante des standards visés ni de processus de soutien lorsqu'il y a des erreurs de compréhension ? Ce qui semble aussi commun des résultats rapportés dans les articles sur l'auto-évaluation est qu'il y a une tendance à ce que les scores attribués par l'auto-évaluation soient surestimés comparativement aux scores attribués par des professeurs, tout au moins dans les premières séances d'auto-évaluation (Kolm, 1984; Gordon, 1991; Khan, 2001; Biernat, 2003).

Norton (2004) rapporte l'expérience où les critères traditionnellement établis pour évaluer sont utilisés comme objectifs d'apprentissage et trouve que les étudiants comprennent mieux les buts qu'ils devraient atteindre. Dans le même sens, les écrits sur

l'évaluation authentique proposent d'utiliser « l'échelle descriptive globale » (terme utilisé par Scallon, 2004, pour traduire le terme anglais "*rubrics*") pour aider l'étudiant à mieux comprendre des attentes dans le développement sa capacité à l'auto-évaluation (Stevens, 2005; Wiggins, 1998; Scallon, 2004; Huba & Freed, 2000; Pangaro, 2000). Par contre, aucune étude n'a été rapportée pour savoir si l'étudiant peut véritablement apprécier la qualité de sa propre performance avec des problèmes complexes dans un contexte de vie réelle, à l'aide de ce type d'échelle.

Seul Pangaro a rapporté une appréciation positive des étudiants avec ce type d'échelle de mesure (Pangaro, 1999), mais aucune mention de l'auto-évaluation n'a été rapportée.

Pour favoriser le développement des compétences cliniques chez les résidents en médecine interne, nous avons créé la méthode des Évaluations formatives des performances cliniques (ÉFPC) (Nguyen, 2006b). Cette méthode d'évaluation formative est constituée de trois composantes centrales : l'évaluation dans un contexte authentique en se basant sur des performances complexes; l'utilisation des professeurs experts pour évaluer formativement des performances des résidents avec une rétroaction approfondie et; l'auto-évaluation.

Notre recherche vise à vérifier l'hypothèse qu'une méthode d'évaluation authentique basée sur des performances dans une perspective d'évaluation formative avec l'utilisation des échelles globales descriptives, favorise le développement de la capacité à l'auto-évaluation. À cette fin, une procédure en a été élaborée :

1. L'étape de la co-évaluation (terme utilisé par Allal, 1999 et Laurier, 2005 pour désigner le processus où simultanément, le professeur et l'étudiant évaluent les performances de l'étudiant selon des critères connus d'avance). Ainsi, à la fin de l'entrevue clinique, chaque résident s'auto-évalue en remplissant la grille

d'observation, qui est la même que celle utilisée par le professeur pour coter les performances du résident.

2. L'étape de rétroaction basée sur les résultats de la co-évaluation où l'étudiant et le professeur comparent leurs grilles. La rétroaction sous forme de discussion interactive porte surtout sur les points de divergence concernant la performance de l'étudiant.

Les questions de recherche sont les suivantes :

1. Quelle est l'évolution de la corrélation entre les scores accordés par les professeurs et les étudiants au cours d'une période de six mois d'expérimentation ?
2. Sur le plan des annotations écrites, y-a t-il une correspondance entre ce qui est noté par les résidents et ce qui est noté par les professeurs ?
3. Sur les stratégies de régulation, quelles sont les stratégies proposées par les étudiants pour s'améliorer ?
4. Quelle est la perception des résidents de la méthode d'ÉFPC pour développer l'auto-évaluation ? La métacognition ?

II. Méthodologie

II.1 Description de la méthode d'ÉFPC

La méthode d'évaluation formative ÉFPC a été développée pour structurer de façon formelle l'évaluation formative dans l'enseignement clinique. Sa description

détaillée a été rapportée dans un article où la validation du contenu et l'établissement des standards de mesure ont été effectuées (Nguyen, 2006b).

Brièvement, il s'agit d'une méthode d'évaluation authentique avec des sessions qui consistent pour chaque résident à interroger, examiner, et répondre aux questions du patient, en se limitant à son problème principal, sous observation directe du professeur qui est aussi le médecin traitant du patient. Par la suite, en l'absence du patient, le résident présente synthétiquement le problème clinique et en tenant compte de toutes les données cliniques et sociales du patient, il propose un diagnostic différentiel, un plan d'investigation et de thérapie pertinents. À la fin de cette étape, de façon indépendante, le résident et le professeur remplissent la grille d'appréciation. La grille d'évaluation est construite sous forme d'échelle globale descriptive à quatre niveaux de standards de performance : 1 étant inadéquat; 2 étant adéquat mais des améliorations sont nécessaires; 3 étant adéquat et est considéré comme le niveau minimal de compétence, et 4 appelé excellent et représente le niveau de professionnel compétent. Les niveaux 1 et 2.

sont considérés comme « novice ». Suite à la comparaison des résultats des deux grilles, commencera la rétroaction qui porte principalement sur les points où il y a une divergence d'opinions.

II.2 Le déroulement du protocole de recherche

Chaque mois, pendant six mois (de février 2005 à juillet 2005), treize résidents ont eu une session d'ÉFPC. Six professeurs ont participé au projet de recherche. Ils ont été assignés aux étudiants par tirage au sort. La date, l'heure et l'endroit des sessions ont été par la suite décidés par le professeur et le résident.

À la fin des sessions d'ÉFPC, les deux grilles ont été acheminées vers le secrétariat. Les professeurs et les résidents n'ont pas eu accès aux performances antérieures des résidents.

À la fin de la période de six mois, une rencontre individuelle avec chaque résident a eu lieu en entrevue semi-structurée enregistrée pour recueillir sa perception sur la méthode d'ÉFPC.

III. Le recrutement des résidents, des professeurs et des patients

Le protocole de recherche décrit dans un autre article où les résultats de la progression des résidents après six mois de suivi ont été rapportés (Nguyen, 2006b) a été utilisé pour recruter résidents, professeurs et patients. Mentionnons que les résidents ont été recrutés par une lettre circulaire envoyée à tous les résidents du programme de Médecine Interne à l'Université de Montréal. Treize ont répondu à l'invitation et se sont engagés à participer à ce projet de recherche étalé sur six mois selon le protocole de recherche pré-établi. Les professeurs ont été recrutés verbalement sur recommandation du Comité de programme de Médecine Interne. Les critères de sélection ont été basés sur leur réputation de dévouement à l'enseignement, leur expérience de l'observation directe, leur connaissance du programme de Médecine Interne ainsi que leur expertise clinique. Dix professeurs cliniciens ont été sollicités et cinq ont décliné l'invitation faute de temps. Le chercheur principal a dû faire partie de l'équipe des professeurs pour maintenir le nombre de sessions d'ÉFPC à un niveau acceptable.

IV. Plan d'analyse des résultats

L'analyse des résultats a été effectuée en quatre parties. La première rapporte l'étude des corrélations entre les scores annotés par des résidents et ceux des professeurs avec le coefficient tau de Kendall pour des valeurs non-paramétriques. En deuxième partie, l'analyse descriptive des fréquences de performances à chaque période a été réalisée pour comparer les résultats entre les scores annotés par les professeurs et ceux par des résidents. Par la suite, en troisième lieu, des annotations par les résidents sur leurs forces et leurs faiblesses ont été analysées qualitativement et mises en relation avec celle des professeurs en utilisant le logiciel Atlas.ti (V5). Finalement, la perception des résidents sur la méthode et l'apport de cette méthode sur leur capacité à s'auto-évaluer sont aussi rapportés par une analyse qualitative effectuée avec le même logiciel.

V. Analyse des résultats

V.1 Résultats de l'évolution des corrélations

Le Tableau VIII rapporte les performances des résidents telles qu'elles sont perçues par les résidents et par les professeurs. Les moyennes des valeurs de performance de la période 1 et 2 ont été calculées pour représenter les valeurs du début. Les moyennes des valeurs des périodes 4 et 6 représentent les valeurs de la fin. Les résultats de la période 5 ont été omis puisque la moitié de ses sessions n'ont pas eu lieu. La compétence clinique globale fut obtenue par l'addition des scores des neuf sous-compétences.

Ces résultats montrent que pour les neuf sous-compétences, à l'exception de celles concernant la proposition d'un plan d'investigation et de thérapie pertinents, la corrélation entre la perception des résidents et celle des professeurs s'améliore à la fin des six mois d'expérimentation. Les sous-compétences où il y a un changement

statistiquement significatif sont celles du questionnaire, de la présentation de cas et de la proposition d'un diagnostic différentiel pertinent. Quant aux sous-compétences que sont l'examen physique et la discussion avec le patient où il y a une corrélation significative dès le début, on observe que cette corrélation reste positive et que dans le cas de l'examen physique, elle est significative à la fin. Pour d'autres sous-compétences telles que le souci de l'impact psychosocial de la maladie et la technique d'entrevue, la corrélation entre la perception des résidents et des professeurs s'améliore, passant respectivement de 0,24 à 0,38 et de -0,02 à 0,32. Par contre, dans le cas des sous-compétences ayant trait à la proposition d'un plan d'investigation et d'un plan thérapeutique pertinent, la corrélation de 0,37 et de 0,36 au début a diminué à 0,08 et 0,02 à la fin.

Quant à la compétence clinique globale qui représente la somme des scores obtenus aux neuf sous-compétences, la corrélation a passé de 0,039 au début à 0,9 à la fin. Ce changement est non seulement statistiquement significatif mais très intéressant pour soutenir notre hypothèse de recherche.

La figure 12 résume ces données et révèle une tendance vers une corrélation positive entre la perception des résidents et des professeurs qui se maintient jusqu'à la fin, et qui, dans certaines sous-compétences, atteint un niveau statistiquement significatif.

Le tableau IX détaille les fréquences des scores que des résidents et des professeurs ont attribués aux performances des résidents. De nouveau, les fréquences des scores du début représentent la moyenne des fréquences de scores des périodes 1 et 2; et ceux de la fin représentent la moyenne des fréquences de scores des périodes 4 et 6.

Au début, les résidents qui s'auto-évaluent ont une nette tendance à se situer vers le centre dans les neuf sous-compétences. Leurs scores d'auto-évaluation se situent de façon prédominante au niveau 2 ou 3. Seules les sous-compétences 2 (souci de l'impact

psychosocial) et 5 (l'examen physique) enregistrent un score du niveau 1 chez deux résidents. Ceci contraste avec l'attribution des scores par les professeurs dont la distribution est plus étalée sur les quatre niveaux de performance.

Les résidents perçoivent de plus dans une grande proportion leurs performances comme adéquates dès le début de l'expérience. Selon eux, au cours des deux premiers mois, à l'exception des sous-compétences 2 (le souci de l'impact social) et 5 (l'examen physique), 65 % de leurs performances sont considérées comme adéquates et notamment situées aux niveaux 3 et 4 avec un intervalle s'étalant de 65 % à 85 %. Par opposition, les professeurs jugent adéquates seulement 45 % des performances avec un intervalle qui s'étale de 46 % à 81 %. Deux exceptions sont les compétences 4 (discussion avec des patients) et 6 (présentation de cas clinique) où 78% des performances ont été perçues par les professeurs comme ayant un niveau « adéquat » dès le début.

Cette discordance de perception entre les résidents et les professeurs s'estompe à la fin, où une meilleure concordance peut être constatée. Ainsi, pour sept des neuf sous-compétences (1, 2, 4, 5, 6, 8, et 9) les résidents et les professeurs ont situé, dans la même proportion et au même niveau de performance : les sous-compétences concernant le questionnaire, la discussion avec les patients, la présentation de cas cliniques, le plan d'investigation et le plan thérapeutique qui ont atteint à la fin le niveau « adéquat » dans plus de 60 % des cas. Toujours selon les deux groupes, les sous-compétences en regard au souci de l'impact psychosocial de la maladie et à l'examen physique restent au niveau 2 (« à améliorer ») dans la même proportion respective d'environ 30 et 45 %.

Par contre, dans les sous-compétences 3 (technique d'entrevue) et 7 (diagnostic différentiel) les résidents se sous-estiment à la fin par comparaison avec l'évaluation des professeurs qui leur ont donné un meilleur score.

Les domaines où il y a une grande discordance entre les résidents (qui se jugent « adéquat ») et les professeurs (qui ont jugé « inadéquat » ou « amélioration nécessaire ») se sont nettement améliorés: Il s'agit des sous-compétences 2 et 4. En effet, quand il s'agit du souci des impacts psychosociaux; au début, 35 % des performances sont cotés par les résidents comme étant aux niveaux 1 et 2 tandis que les professeurs les jugent plutôt à 56%. À la fin, les deux se trouvent à la même proportion; car entre 34 et 37 % des performances sont jugées comme étant situé au niveau 2. Concernant le questionnaire des patients, 27% des performances sont considérées par les résidents comme ayant besoin d'améliorations contre 42 % selon l'avis des professeurs. Ces proportions sont passées à 92 % et 82 % à la fin où tous les deux pensent que la performance a atteint au moins le niveau adéquat.

Par ailleurs, sur le plan strictement auto-évaluatif des performances, dans deux sous-compétences où les résidents s'accordent un niveau élevé de performance au début, il y a une diminution des scores auto-attribués. Il s'agit de la sous-compétence 4 (« discussion avec les patients ») et 9 (« plan thérapeutique »). Les sept autres sous-compétences, les performances du début se sont améliorées par rapport à la fin avec un étalement sur les deux niveaux dans les sept sous-compétences; la majorité se situant au niveau 3 (de 46 % à 79 %) et une minorité au niveau 4 (de 4 % à 23 %).

Le tableau X résume l'évolution de la proportion des concordances des scores attribués par les résidents et par des professeurs. Cette proportion est calculée comme la somme des fréquences où les deux ont attribué les mêmes scores. Dans l'ensemble, il y a une tendance à l'amélioration des proportions de concordance entre le début et la fin. Ces résultats sont aussi illustrés par la figure 13. Il y a des limites à cette méthode de calcul mais nous y reviendrons dans la discussion.

V.2 Analyse qualitative des annotations

Ce rapprochement de jugement entre les professeurs et les résidents est aussi reflété dans les annotations écrites. L'analyse de ces annotations écrites s'est effectuée par la technique d'encodage avec le logiciel ATLAS.ti (V5.0). Deux familles de codes représentant des annotations «à améliorer» et « à maintenir » ont été créées. Une troisième famille de codes pour l'ensemble des suggestions pédagogiques a aussi été créée. Le tableau XI rapporte les résultats qualitatifs de cette analyse. Au total, il y a 158 annotations de professeurs et 69 de résidents. Un chevauchement entre les préoccupations des professeurs et des résidents peut être constaté, notamment en ce qui concerne l'examen physique, le questionnaire et le diagnostic différentiel. Les trois occupent respectivement jusqu'à 70 % et 88 % de commentaires des professeurs et des résidents portant sur les domaines à améliorer. Par contre, les professeurs apportent une plus grande attention que les résidents à certains domaines où les dimensions relationnelles ou de structure sont prédominantes. Il s'agit de la présentation de cas, du souci de l'impact psychosocial de la maladie et du contact avec les patients.

Seulement 8 % des commentaires des résidents font allusion à des améliorations nécessaires dans les techniques d'entrevue (4 %) et le souci de l'impact psychosocial de la maladie sur la vie du patient (4 %). Aucune mention ne touche la qualité de leur présentation de cas. Cette sous-compétence est pourtant le sujet de 17 % des commentaires des professeurs sur les faiblesses des résidents, tandis que 8 % des commentaires des professeurs traitent de la question « impacts psychosociaux de la maladie sur la vie des patients ».

Les grilles font ressortir au sujet des forces, qu'à la fin, les résidents et les professeurs s'accordent à dire que le questionnaire est adéquat. Mais on retrouve de nouveau ici une plus grande préoccupation chez les professeurs pour l'aspect « relation patient- médecin » que chez les résidents.

Les exemples de commentaires qui se trouvent au tableau XII sont des exemples qui illustrent les commentaires retrouvés sur les grilles d'ÉFPC dans les quatre domaines de compétences qui préoccupent communément les résidents et les professeurs. Essentiellement, les commentaires se recoupent mais ceux des professeurs sont beaucoup plus élaborés, ceux des résidents semblent plus laconiques et moins détaillés.

Finalement, aux questions sur les activités pédagogiques à faire ou qui ont été faites pour s'améliorer, les résidents ne font au total que 14 commentaires et la majorité (7/14) mentionne la nécessité d'une lecture plus approfondie. Trois commentaires ont porté sur l'effort de faire un questionnaire en y incluant les impacts psychosociaux de la maladie. Un seul commentaire a porté sur l'examen physique et l'effort d'avoir du "feedback" d'autres professeurs sur leur technique de l'examen physique.

Les suggestions pédagogiques des professeurs sont en revanche plus nombreuses (39) et légèrement plus diversifiées. De nouveau, la nécessité de la lecture reste la recommandation la plus fréquemment suggérée (50 %), que ce soit au sujet de l'examen physique spécifique ou des connaissances ciblées qui permettent de réaliser un questionnaire pertinent et un meilleur diagnostic différentiel.

V.3. Analyse qualitative de la perception des résidents

Lors des entrevues individuelles avec les résidents, deux des questions portaient sur l'auto-évaluation. La première question pour vérifier la perception des résidents dans le développement de la compétence à s'auto-évaluer et la deuxième, sur leur développement à la métacognition :

« ... est-ce que cette méthode vous permet de mieux vous évaluer vous-mêmes, de mieux vous connaître vous-mêmes ? »

« ... est-ce que cette méthode permet de mieux diriger vos études ? ... »

Au total, des 246 commentaires codés, 47 portent sur l'auto-évaluation.

Sur 13 résidents, 3 n'ont pas trouvé que leur capacité de s'auto-évaluer avait changé :

« ... auto-évaluation a peu changé ... »

« ... je ne pense pas que ça a changé ... »

Les autres reconnaissent que les ÉFPC leur permettent de :

« ... mieux réaliser mes faiblesses »

et de réaliser que :

« ... souvent, je me sous-estimais alors que mon patron m'estimait ... »

« ... cela a aidé à développer de la confiance ... »

et

« ... je vois moi-même un changement ... »

Par contre, à la deuxième question sur la régulation, aucun des résidents n'a spontanément reconnu un changement dans ses habitudes de travail. Plusieurs n'ont même pas, de façon consciente, réalisé qu'ils ont tenté de lire pour combler les déficiences remarquées par leurs professeurs.

De plus, 20 commentaires des résidents sur 246 commentaires portent sur la grille utilisée pour recueillir des données sur leur performance. Les 13 sont d'accord à reconnaître que :

« La grille donne un bon aperçu ... » de ce qui est attendu.

Nous avons pu analyser 110 des commentaires portant sur la rétroaction.

Les résidents ont unanimement bien apprécié les rétroactions qui a eu lieu après l'auto-évaluation par une discussion interactive parce qu'elle était :

« ... constructive ... »,

« ... sérieuse ... », et

« ... dans 80 % des cas juste ... »

Même si dans certains cas, il y a eu :

« ... étirement un peu trop... », ou

« ... exagération de certains détails ... »

Les opinions sont partagées sur l'organisation des ÉFPC. À la question portant sur leur perception de la méthode, les 13 résidents ont aussi unanimement répondu avoir apprécié le temps qu'ils ont considéré comme « ... privilégié... » où ils ont eu un professeur « ... qui prend le temps de [nous] regarder ... ».

L'appréciation globale positive se reflète dans 64 des commentaires sous forme de :

« ... méthode d'évaluation juste ... »

« ... excellente méthode ... »

« ... beaucoup moins stressante que des ECOS ... »

Par contre, 60 commentaires sont négatifs, la grande majorité (32) portant sur l'organisation. Ainsi, un résident trouve « ... le temps limitatif... » et d'autres aimeraient avoir « ... accès au dossier du patient... » et « ... aux livres, comme dans la vraie vie... »

VI. Discussion et conclusion

Les ÉFPC se sont avérées utiles comme méthode pour améliorer la capacité des résidents à s'auto-évaluer. L'analyse quantitative des résultats d'auto-évaluation avec les ÉFPC montre une amélioration significative de la capacité des résidents à s'auto-évaluer dans cinq des neuf sous-compétences, notamment le questionnaire, le souci des impacts psychosociaux de la maladie sur la vie des patients, la technique d'entrevue, la présentation de cas et le diagnostic différentiel. Pour les deux autres sous-compétences concernant l'examen physique et la discussion avec le patient, les corrélations positives significatives demeurent stables. Deux sous-compétences (plan d'investigation et plan thérapeutique) n'ont par contre pas montré de résultats ayant une différence statistiquement significative. Dans ces deux derniers cas, l'analyse des proportions montre cependant qu'une proportion de 90% de concordance entre les scores des résidents et ceux des professeurs. Il est de plus rassurant que dans aucun cas, nous ne retrouvons des résidents auxquels les professeurs ont donné un score du niveau 1 et qui seraient attribués des scores 3 ou 4. Comme nous pouvons le constater, les résidents ont réussi à se percevoir eux-même comme leur professeur les ont perçus après six mois de formation. Des neuf sous-compétences, quatre ont un niveau de corrélation qui se situe au-dessus de 0,5.

Le rôle de l'évaluation simultanée par les professeurs et la rétroaction qui en découle a une importance significative. Ainsi, lorsqu'on observe les domaines où les résidents et les professeurs s'accordent à dire que des améliorations sont nécessaires, le coefficient des scores à la fin atteint un niveau statistiquement significatif (notamment le questionnaire et l'examen physique).

Quant aux domaines qui font l'objet de regards moins critiques, que ce soit par les professeurs ou par les résidents, le rapprochement des deux perceptions reste moins

important avec un coefficient de corrélation qui oscille autour de 0,3 (la technique d'entrevue, le souci des impacts psychosociaux et la discussion avec les patients).

Dans les cas où aucune attention n'a été accordée, soit par les résidents, soit par les professeurs, en se basant sur leur perceptions annotées de leurs forces et de leurs faiblesses (i.e. : l'investigation et le traitement), peu de progrès est noté et la corrélation entre la perception des résidents et celle des professeurs devient non significative à la fin.

Globalement, l'appréciation des résidents de leur propre compétence clinique est en bonne corrélation avec l'appréciation des professeurs à la fin de l'expérimentation, avec un tau de Kendall de 0,9. Cela se reflète aussi dans les commentaires des résidents qui ont unanimement trouvé la méthode utile et, cette méthode leur a permis de progresser. Des résidents constatent même en majorité leur propre amélioration et qu'ils ont une meilleure connaissance de leurs forces et de leurs faiblesses.

De plus, même si les données ne sont recueillies que par un professeur pour chaque session, il semble que lorsque les performances se répètent sous observation directe par différents professeurs, les résidents considèrent leurs mesures comme fiables. Ainsi, les résidents ont trouvé les ÉFPC :

« ... utiles parce qu'à force de se faire dire deux, trois fois par différents patrons, on finit par le croire et changer ... »

Quant au développement de la métacognition, il demeure encore très modeste. Le processus de régulation proposé par des résidents se limite souvent à la lecture personnelle; ce qui correspond aussi à l'opinion des professeurs qui estiment qu'il faut adopter la même stratégie pour s'améliorer.

Ces données vont au-delà des données disponibles publiées en auto-évaluation en formation clinique où les corrélations entre les étudiants et les professeurs se situent en

moyenne à 0,39 (Ward, 2002). Par la suite, Ward (2003) a publié des données sur l'auto-évaluation des résidents sur leur technique opérative et obtient un niveau de corrélation entre les résidents et leur professeur qui a passé de 0.5 à 0.63 après qu'ils ont visionné l'enregistrement audiovisuel de leur performance, mais seulement dans le domaine psychomoteur. Calhoun (1990) et Arnold (1985) qui ont étudié longitudinalement la capacité à l'auto-évaluation n'ont pas réussi à démontrer une amélioration. Leurs recherches n'ont pas mentionné s'il y a eu une rétroaction après l'auto-évaluation.

Nos données nous ont permis d'observer que sept des neuf compétences atteignent un niveau de corrélation qui dépasse 0,3 à la fin de l'expérimentation. Il est de plus rassurant que la fréquence avec laquelle les résidents et les professeurs pensent que des améliorations sont nécessaires est la même à deux sous-compétences qui n'ont pas atteint ce niveau de corrélation. L'évaluation globale des compétences atteint une corrélation de 0,9 , ce qui, à notre connaissance, n'a jamais été rapporté auparavant.

Nos données montrent par ailleurs la même tendance observée dans l'auto-évaluation. Les résidents ont tendance à surestimer leur performance (Gordon, 1991; Schmidt, 2001; Matheos, 2004). C'est ce qui est reflété au début. La tendance vers la cote au centre de l'échelle d'auto-évaluation des résidents au début de l'expérience montre de plus leur tendance à s'évaluer globalement. Cela va dans le même sens de constatations de Gordon (1991). Des changements ont été constatés dans la capacité d'un résident à l'auto-évaluation lors de l'utilisation de la grille d'appréciation globale descriptive combinant avec une rétroaction approfondie de façon répétée.

Finalement, notre étude semble confirmer le rôle important de la rétroaction individualisée dans le développement des compétences. Comme cette rétroaction a eu surtout le format de discussion interactive, lorsque les deux acteurs s'entendent sur les sous-compétences à améliorer, ce n'est pas seulement la performance des résidents qui

s'améliore, mais aussi sa capacité de s'évaluer sur la même sous-compétence qui s'améliore. De plus, dans le contexte de l'évaluation authentique, avec une grille d'évaluation globale descriptive, tous les résidents apprécient fortement la combinaison de la co-évaluation et de la rétroaction. Ils connaissent mieux ce qui est attendu et perçoivent qu'ils ont été évalués à leur juste valeur, ce qui renforce leur motivation à s'améliorer; tel que le préconise Laveault (1999).

Les limites de cette étude, qui sont le petit nombre d'étudiants et de professeurs, ainsi que l'absence d'un groupe de contrôle, rendent difficile la généralisation. De plus, le calcul par les proportions de concordance est aussi affecté puisque ceci se repose sur le nombre de jugements des professeurs. Un nombre réduit de professeur diminue la variabilité potentielle des annotations. Des recherches plus poussées, sur une plus longue période de temps, avec des groupes d'étudiants plus importants seraient nécessaires pour vérifier si l'amélioration de la capacité d'auto-évaluation est soutenue et à quel moment le rôle de la rétroaction serait moins important puisque les résidents se sont appropriés des critères d'évaluation et des standards à atteindre.

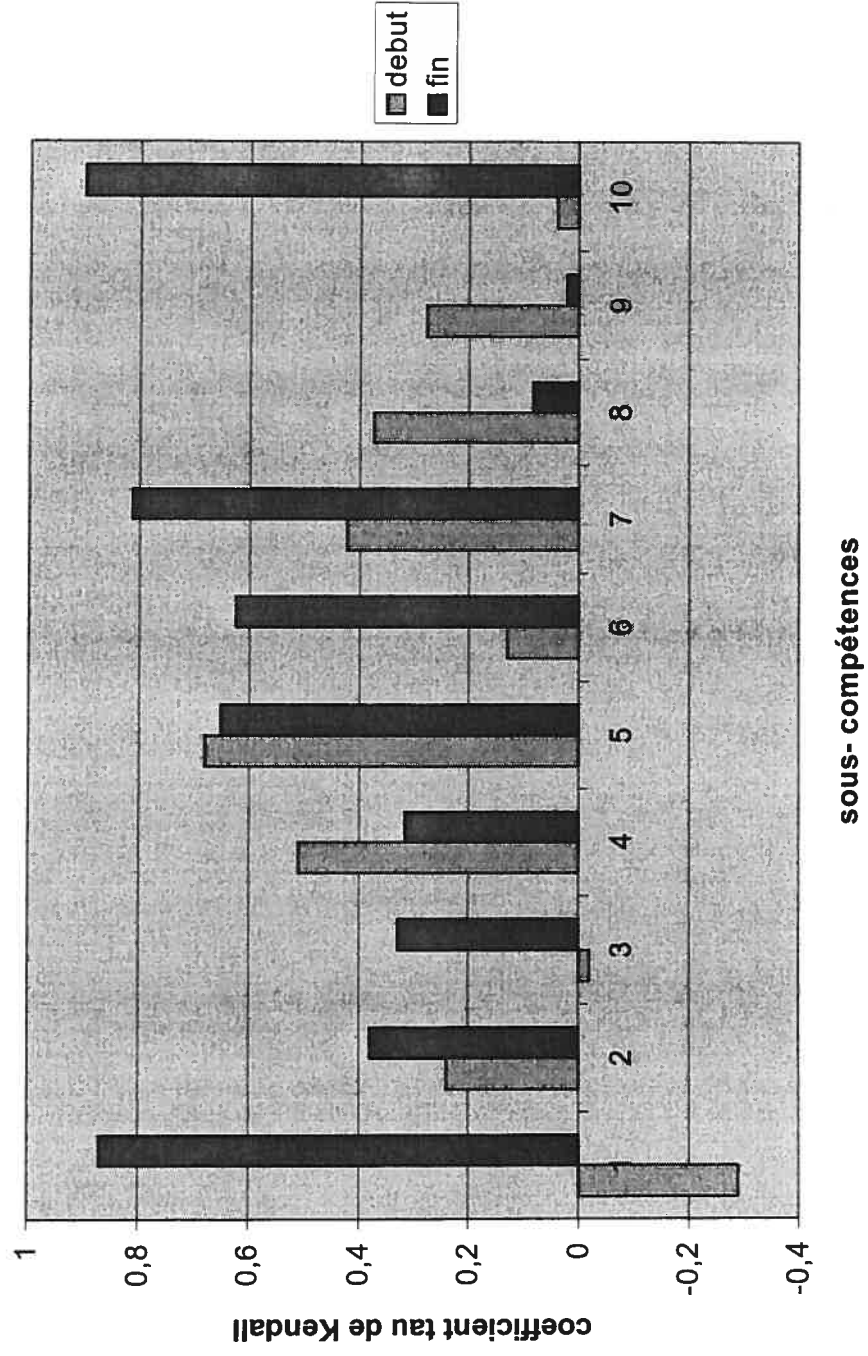
TABLEAU VIII : CORRÉLATION DES PERFORMANCES AU DÉBUT ET À LA FIN

	Début			Fin		
	Résident moyenne \pm écart type	Professeur moyenne \pm écart type	Corrélation coefficient tau de Kendall	Résidents moyenne \pm écart type	Professeurs moyenne \pm écart type	Corrélation coefficient tau de Kendall
Sous-compétence 1 : Questionnaire	2.73 \pm 0.33	2.58 \pm 0.49	-0.29	3.08 \pm 0.34	3.08 \pm 0.45	0.87(*)
Sous-compétence 2 : Souci des impacts psychosociaux	2.77 + 0.60	2.26 + 0.56	0.242	3.08 + 0.34	2.77 + 0.73	0.38
Sous-compétence 3 : Technique d'entrevue	2.81 + 0.43	2.96 + 0.38	-0.02	3.30 + 0.38	3.42 + 0.53	0.329
Sous-compétence 4 : Discussion avec le patient	2.96 + 0.58	2.11 + 0.68	0.51(*)	3.15 + 0.52	3.19 + 0.32	0.315
Sous-compétence 5 : Examen physique	2.5 + 0.58	2.12 + 0.68	0.68(*)	2.62 + 0.58	2.69 + 0.59	0.651(*)
Sous-compétence 6 : Présentation de cas	2.66 + 0.47	2.42 + 0.45	0.138	2.65 + 0.45	2.58 + 0.45	0.625(*)
Sous-compétence 7 : Diagnostic différentiel	2.65 + 0.42	2.42 + 0.53	0.423	2.65 + 0.37	2.57 + 0.64	0.813(*)
Sous-compétence 8 : Investigation	2.69 + 0.59	2.50 + 0.65	0.374	2.88 + 1.42	2.88 + 0.30	0.083
Sous-compétence 9 : Plan thérapeutique	2.69 + 0.38	2.62 + 0.60	0.359	2.76 + 0.26	2.76 + 0.44	0.021
Compétence clinique globale 10 :	24.23 + 3.13	21.81 + 2.70	0.039	26.0 + 2.90	25.96 + 2.80	0.90(*)

(*) $p < 0.05$

Figure 12 : Évolution de la corrélation des performances

Figure 12: Evolution de la corrélation des performances



TABEAU IX : FRÉQUENCE DES PERFORMANCES SELON LES PROFESSEURS ET LES RÉSIDENTS

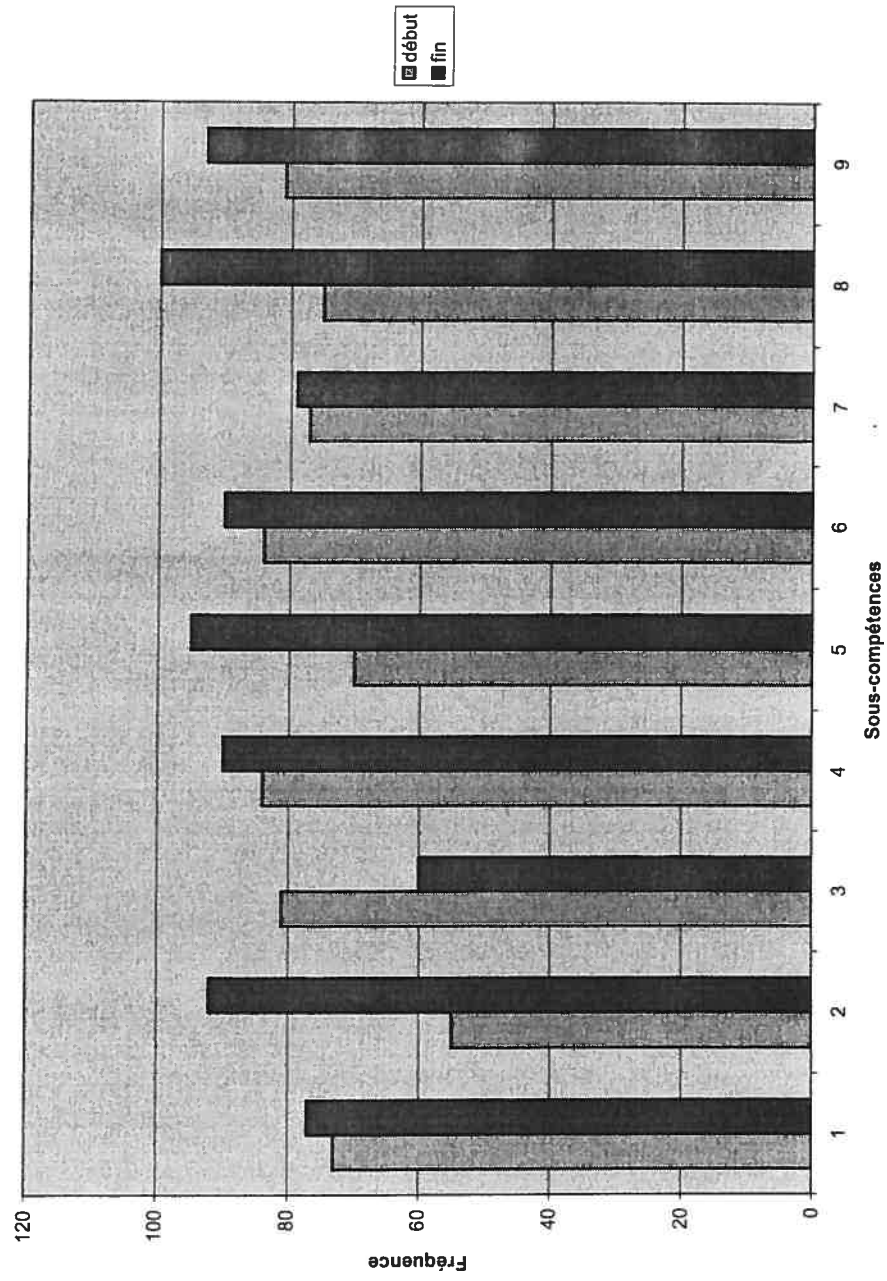
		Début (%)										Fin (%)									
		Performance (niveau) Résident					Performance (niveau) Professeur					Performance (niveau) Résident					Performance (niveau) Professeur				
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Sous-compétence 1 : QUESTIONNAIRE	0	27	73	0	0	0	42	46	7	0	5	79	16	0	18	61	21				
Sous-compétence 2 : SOUCI DES IMPACTS PSYCHO- SOCIAUX	4	31	62	39	4	4	52	8	12	0	34	55	13	5	32	47	19				
Sous-compétence 3 : TECHNIQUE D'ENTREVUE	0	23	69	8	0	0	19	54	27	0	32	69	10	0	5	45	50				
Sous-compétence 4 : DISCUSSION AVEC LES PATIENTS	0	15	69	15	0	0	22	78	0	5	5	67	23	0	0	77	23				
Sous-compétence 5 : EXAMEN PHYSIQUE	8	46	35	12	30	30	37	21	4	0	46	46	8	0	50	41	9				
Sous-compétence 6 : PRÉSENTATION DE CAS	0	15	69	15	0	0	22	78	0	5	5	67	23	0	0	77	23				
Sous-compétence 7 : DIAGNOSTIC DIFFÉRENTIEL	0	35	65	0	4	4	46	42	4	0	38	62	0	5	42	41	13				
Sous-compétence 8 : INVESTIGATION	4	13	83	0	8	8	29	58	4	0	18	74	8	0	18	74	8				
Sous-compétence 9 : PLAN THÉRAPEUTIQUE	0	29	71	0	18	18	15	63	4	0	24	72	4	0	31	65	5				

TABEAU X : PROPORTION DE CONCORDANCE DES SCORES

	Début (%)	Fin (%)
Sous-compétence 1	73	77
Sous-compétence 2	55	92
Sous-compétence 3	81	60
Sous-compétence 4	84	90
Sous-compétence 5	70	95
Sous-compétence 6	84	90
Sous-compétence 7	77	79
Sous-compétence 8	75	100
Sous-compétence 9	81	93

Figure 13 : Proportion de la concordance des scores

Figure13 :proportion des concordance des scores



TABEAU XI : ANNOTATIONS ET COMMENTAIRES SUR LA GRILLE PAR RÉSIDENTS ET PROFESSEURS

Perception	Professeurs n =158 commentaires (%)	Résidents n = 69 commentaires (%)
1. À améliorer	113 commentaires / 178 (63.5%)	52 / 69 commentaires (75 %)
	- L'examen physique 32 (31)	- L'examen physique 25(48)
	- Questionnaire 22 (21)	- Questionnaire 15 (28)
	- Diagnostic différentiel 19 (18)	- Diagnostic différentiel 6 (12)
	- Présentation de cas 18 (17)	- Impact psychosocial 2 (12)
	- Impact psychosocial 7 (7)	- Technique d'entrevue 2 (4)
	- Investigation 3 (3)	- Traitement 2(4)
	- Traitement 2 (2)	
	- Discussion 1 (1)	
	- Total : 104 (100%)	- Total : 52 (100 %)
2. À maintenir	65 / 178 commentaires (36.5 %)	17 / 69 commentaires (25 %)
	- Questionnaire 21(39)	- Technique d'entrevue 6(35)
	- Contact avec le patient 15 (28)	- Questionnaire 4(24)
	- Examen physique 12 (22)	- Discussion avec les patients 3 (18)
	- Technique d'entrevue 3 (6)	- Diagnostic différentiel 3 (18)
	- Impact psychosocial 1 (2)	- Examen physique 1 (5)
	- Investigation 2(3)	
	- Total : 54 (100%)	- Total : 17 (100%)

Tableau XII : Exemples des commentaires des résidents et des professeurs sur les forces et les faiblesses des résidents

Résidents	Professeurs
<p>I. L'examen physique</p> <p>« ...L'examen physique structuré et raffiné ... »</p> <p>« ...structurer l'examen rhumatologique pour ne rien oublier ... »</p>	<p>« ... examen physique pertinent à améliorer ... »</p> <p>« ... examen physique à faire avec plus de rigueur ... »</p>
<p>II Le questionnaire</p> <p>« ... questionnaire plus succinct ... »</p> <p>« ... questionnaire complet ...t »</p>	<p>« ... attention à ne pas perdre trop de temps lors du questionnaire ... »</p> <p>« ... bon questionnaire, les données sont bien intégrées ... »</p>
<p>III Diagnostic différentiel</p> <p>« ... élargir un peu le diagnostic différentiel ... »</p> <p>« ... intégrer les données cliniques au diagnostic différentiel ... »</p>	<p>« ... bien argumenter le diagnostic différentiel ... »</p> <p>« ... le diagnostic selon le plus probable au moins probable ... »</p>
<p>IV Impact psychosocial</p> <p>« ...questionner davantage l'impact de la maladie ... »</p> <p>« ... impact de la maladie pour les patients ... »</p>	<p>« ... impact du traitement et de l'investigation sur le patient ... »</p> <p>« ...se préoccuper beaucoup de l'impact de la maladie sur la vie du patient ... »</p>

RÉFÉRENCES

ACGME (1999). *ACGME Outcome Project, General competencies* : <http://www.acgme.org/outcome/comp/compfull.asp>, accédé le 28 février 2006.

Allal, L. (1999). Impliquer l'apprenant dans le processus d'évaluation : promesses et pièges de l'autoévaluation. In C. Depover & B. Noël (dir.), *L'évaluation des compétences et des processus cognitifs; modèles pratiques et contextes*. Bruxelles : De Boeck Université.

Arnold, L., Willoughby T.L., Calkins E.V. (1985). Self-evaluation in undergraduate medical education : a longitudinal perspective. *Journal of Medical Education*, 100 (1) : 21-28.

Battistone, M. J., Mine C., Sande M. A., Pangaro L. N., Hemmer P. A., Shomaker T. S. (2002). The feasibility and acceptability of implementing formal evaluation sessions and using descriptive vocabulary to assess student performance on a clinical clerkship. *Teaching & Learning in Medicine*, 14 (1) : 5-10 .

Biernat, K.; Simpson, D., Duthie, E. : Dawn, B.; London, R. (2003) Primary care residents self-assessment skills in dementia. *Advances in Health Sciences Education*, 8 : 105-110.

Blackstein, D., Agnidis Z., Regehr R., Reznick R. (2004). The effectiveness of video feedback in the acquisition of orthopedic technical skills. *American Journal of Surgery*, 187 : 427-432.

Boud, D. (2000). Sustainable assessment : rethinking assessment for the learning society. *Studies in Continuing Education*, 22 (2) : 151-167.

Calhoun, J. G.; Ten Haken J. D.& Woolliscroft J. O. (1990). Medical students development of self- and peer-assessment skills : a longitudinal skill. *Teaching & Learning in medicine*, 2 (1) : 25-29.

Falchikov, N.; Boud, D. (1989). Students self-assessment in Higher Education : a meta-analysis. *Review of Educational Research*, 59 (4) : 395-430.

Fallows, S., Chandramohan B. (2001). Multiple approaches to assessment : reflections on use of tutor, peer, and self-assessment. *Teaching in Higher Education*, 6 (2) : 229-246.

Frank, K. J. R., Jabbour M., Tugwell P. & al. (1996). Skills for the new millenium : report of the societal needs working group; CanMEDS 2000 Project. *Annals of Royal College of physicians and surgeons of Canada*, 29 : 206-216.

Gagné, R. M., Wager W. W., Grolas K. C., Leller J. M. (2005). *Principles of instructional design*, (5th ed.). New-York : Thomson Wadsworth.

Gordon, M. J. (1991). A review of the validity and accuracy of self-assessment in health professions training. *Academic Medicine*, 66 (12) : 762-769.

Gordon, M. J. (1992). Self-assessment programs and their implications for health professions training. *Academic Medicine*, 67 (10) : 672- 679.

Herbert, W. P.; Mc Gaghie W. C., Droegemueller W., Riddle M., Maxwell K.L. (1990). Student evaluation in obstetric and gynecology : self- versus deparmental assessment. *Obstetrics & Gynecology*, 76 :458-461.

Huba, M. E. , Freed J. E. (2000). *Learner – centered assessment on college campuses, shifting the focus from teaching to learning*. USA : Allyn & Bacon.

Khan, K. S., Davies D. A, Gupta J. K. (2001). Formative self-assessment using multiple true- false questions on the Internet feedback according to confidence about correct knowledge. *Medical Teachre*, 23:158-163.

- Klenowski, V. (1995). Student self-evaluation processes in student – centred teaching and learning contexts of Australia and England. *Assessment in Education*, 2 (2) : 145-163.
- Kolm, P., Verhurst, S.J. (1984). Comparing self and supervisor evaluations : a different view. *Proceedings of Annual Conference on Research in Medical Education*, 23 : 63-67.
- Laurier, M. D., Toussignant R. & Morissette D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages*, (3^e ed.). Québec : Gaëtan Morin.
- Laveault, D., Leblanc R. & Leroux, J. (1999). Autorégulation de l'apprentissage scolaire : interaction entre processus métacognitifs et déterminants de la motivation. In C. Depover & B. Noël (dir.) : *L'évaluation des compétences et des processus cognitifs. Modèle, pratiques et contextes*, 81-98. Bruxelles : De Boeck Université.
- Mattheos, N., Nattestad A., Falf-Nilson E., Attström R. (2004). The interactive examination : assessing students' self-assessment ability. *Medical Education*, 38 : 378-389.
- Miller, P.J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28 (4) : 383-394.
- Nguyen, D.Q. (2006a). Approche par objectifs ou approche par compétences : quelles implications pour l'enseignement et pour l'évaluation dans la formation clinique ? Article qui sera soumis à la revue *Pédagogie médicale*.
- Nguyen, D. Q. (2006b). Modélisation d'une grille d'évaluation formative des performances dans une perspective d'approche par compétences : un exemple de formation post-doctorale en Médecine Interne. Article qui sera soumis à la revue *Mesure & évaluation en éducation*.
- Norton, L. (2004). Using assessment criteria as learning criteria : a case study in psychology. *Assessment & Evaluation in Higher Education*, 29 (6) , 687-702.

Orsmond, P., Merry, S.; Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self - assessment. *Assessment and Evaluation in Higher Education*, 27(4):309-323.

Pangaro, L. (1999). A new vocabulary and other innovations for improving descriptive in-training evaluations. *Academic Medicine*, 74 (11) : 1203-1207.

Pangaro, L. N. (2000). Investing in descriptive evaluation : a vision for the future of assessment. *Medical Teacher*, 22 (5) : 478-481.

Saddler, R. (1998). Formative assessment : revisiting the territory. *Assessment in education*, 5 (1), 77-84.

Scallon, G. (2004). L'évaluation des apprentissages dans une approche par compétences. Québec : Éditions du Renouveau pédagogique.

Schmidt, H. G., Van der Molen H.T. (2001). Self-reported competency ratings of graduates of a problem-based medical curriculum. *Academic Medicine*, 76 : 466-468.

Ward, M., Gruppen L. & Regehr G. (2002). Measuring self-assessment : current state of the art. *Advances in Health Sciences Education*, 7 : 63-80.

Ward, M., Mac Rae H., Schlachta C., Mamazza J., Poulin E., Reznick R., Regehr G. (2003). Resident self-assessment of operative performance. *American Journal of Surgery*, 185 : 521-524.

Wiggins, G. (1998). *Educative assessment : designing assessments to inform and improve student performance*. San-Francisco : Jossey-Bass.

Conclusion générale

Notre thèse a pour but de vérifier l'hypothèse selon laquelle l'évaluation authentique dans une perspective d'évaluation formative au niveau des études supérieures favorise le développement des compétences professionnelles. La nouvelle méthode d'évaluation, appelée les Évaluations Formatives des Performances Cliniques(ÉFPC), a été créée pour prouver cela.

La conception de cette méthode repose sur les observations qui découlent du premier article dont le but principal est de comparer deux approches pédagogiques (par objectifs et par compétences). L'analyse des deux approches pédagogiques qui occupent une place importante dans des curriculums et programmes de formation, à tous les niveaux, a permis de constater certaines différences fondamentales entre les deux approches. La première est caractérisée par la minutie avec laquelle les objectifs pédagogiques sont élaborés et spécifiés. Les difficultés constatées par la suite découlent de la tendance, au moment de la planification des activités d'apprentissage et d'évaluation, de fragmenter les tâches et de mettre l'accent sur les objectifs cognitifs à niveau taxonomique plus faible et de négliger les objectifs cognitifs plus complexes ainsi que des objectifs psychomoteurs et affectifs.

L'approche par compétences quant à elle met l'accent sur le savoir agir et insiste sur l'intégration des savoirs pour résoudre des problèmes complexes et signifiants pour les étudiants. À l'heure actuelle, des données en recherche ne permettent par contre pas de déterminer les meilleures stratégies d'apprentissage et d'évaluation qui assureront le développement des compétences.

Les études très abondantes sur différentes facettes de l'évaluation de l'atteinte des objectifs nous habituent à conceptualiser la fidélité de mesure sous forme classique de réplicabilité et de consistance interne et de la validité sous forme de généralisabilité par différentes méthodes d'échantillonnage des items. Comme nous l'avons évoqué dans l'article 1, l'évaluation avec des items mesurant des dimensions uniques ne semble pas suffisante pour évaluer des compétences. Il ressort que l'évaluation des compétences devrait être basée sur des performances complexes, et ce concept semble être accepté par des partisans de l'approche par compétences.

En adoptant l'évaluation des performances complexes pour évaluer des compétences, dans la formation clinique, il fallait trouver une méthode qui puisse évaluer les performances complexes dans une perspective d'évaluation formative. C'est ainsi que la méthode des ÉFPC a été conçue à l'intention des étudiants de formation spécialisée en Médecine Interne.

En conceptualisant les ÉFPC, il a aussi été important de déterminer quel type d'outil de mesure serait approprié pour recueillir des données nécessaires à l'évaluation formative des compétences cliniques. L'article 2 démontre alors qu'une échelle descriptive globale à 4 niveaux de standards de performance et avec 9 items qui représentent les tâches cliniques professionnelles semble répondre aux besoins d'évaluation formative des compétences cliniques dans un contexte de curriculum entièrement sous forme de stages. Ainsi, dans un contexte où chaque étudiant a un bagage de connaissances différent, évolue à un rythme différent, est encadré par des professeurs avec des compétences professionnelles et pédagogiques variables, et a des expériences d'apprentissage variables et imprévisibles, une méthode d'évaluation structurée et formelle avec un outil de mesure basé sur les compétences professionnelles à développer semble émerger comme une méthode potentiellement utile.

Comme l'évaluation formative est un processus qui survient au cours de la formation et qui a pour but d'aider les étudiants à progresser, notre démarche pour valider l'ÉFPC comme méthode d'évaluation formative oblige alors une étude qui détermine si les étudiants ont véritablement progressé avec les ÉFPC. Dans l'article 3, nos résultats ont alors effectivement montré que globalement, les compétences cliniques des résidents en médecine ont progressé de façon significative au cours de six mois de mise à l'essai. L'analyse quantitative des données semble alors démontrer que six sous-compétences cliniques ont significativement progressé. Il s'agit des sous-compétences concernant le questionnaire du patient, la recherche des impacts psychosociaux de la maladie, l'examen physique, les discussions avec le patient, les techniques d'entrevue et le plan d'investigation. L'analyse qualitative des commentaires des professeurs a permis de réaliser que ces six domaines ont été ciblés par les professeurs lors des périodes de rétroaction, ce qui a permis d'atteindre ce niveau de progression. Dans le cas des trois autres sous-compétences, la progression telle qu'elle est jugée par les professeurs est moins importante en parallèle à une moins grande importance accordée lors de la rétroaction.

Le quatrième article qui est entièrement consacré à l'analyse des résultats du développement de l'auto-évaluation démontre de nouveau le rôle important de la rétroaction. Les données semblent montrer, contrairement aux croyances générales, que l'utilisation d'une grille d'évaluation globale descriptive dans un contexte d'évaluation authentique avec des performances complexes ne suffit pas pour développer la capacité à l'auto-évaluation. Nos données semblent plutôt démontrer que c'est le processus de co-évaluation avec une discussion interactive en se basant sur les points divergents qui permet un meilleur développement de l'auto-évaluation. De nouveau, le rôle de la rétroaction semble être crucial dans le développement de la capacité à l'auto-évaluation.

Cela se reflète dans quatre sous-compétences, et leur atteinte d'un seuil de corrélation significatif entre l'évaluation des professeurs et l'auto-évaluation. Mais, outre la rétroaction, le regard critique de l'étudiant sur lui-même a aussi une répercussion significative sur le développement de sa capacité à s'auto-évaluer. Ainsi, la combinaison de l'analyse qualitative des commentaires des étudiants et des professeurs a permis de comprendre que lorsque les deux se soucient des mêmes domaines, il y a une progression à la fois des performances de l'étudiant et de sa capacité à l'auto-évaluation. Ceci est constaté dans trois domaines de sous-compétences: le souci des impacts psychosociaux, les techniques d'entrevue et les discussion avec le patient. Par contre, dans les domaines où les étudiants ne semblent pas apporter une attention particulière, aucune amélioration de la capacité à l'auto-évaluation n'a pu être constatée. Il s'agit des deux sous-compétences qui concernent l'investigation et le plan thérapeutique.

Sur le plan de l'acceptabilité et de faisabilité, l'analyse qualitative des perceptions, tant des professeurs que des étudiants, permet de constater que cette méthode d'évaluation avec une tâche complète et complexe, mais réaliste, semble être fortement appréciée par des étudiants et par des professeurs. Dans le premier cas, les étudiants ont tous trouvé que la rétroaction a été basée sur des faits réels qui reflètent leur performance véritable. Ils ont de plus considéré comme un moment privilégié cette heure d'activité combinant l'évaluation et l'enseignement individualisé. Les professeurs ont éprouvé de la satisfaction d'avoir pu évaluer un résident de façon complète et globale et faire une rétroaction franche en se basant sur des faits constatés et des standards préalablement établis. Les deux groupes estiment que le facteur limitatif reste encore le temps. L'implantation éventuelle de cette méthode d'évaluation devrait tenir compte de cette contrainte.

Nos données ont aussi révélé que l'auto-évaluation est une compétence qui peut se développer mais qui requiert l'aide des professeurs. L'utilisation d'une grille globale

descriptive dans une approche d'évaluation authentique facilite la tâche au professeur pour guider l'étudiant à mieux comprendre les attentes et à s'approprier des standards d'excellence dans son cheminement pour développer ses capacités à s'auto-évaluer.

En conclusion, avec les ÉFPC, nous avons pu démontrer qu'une méthode d'évaluation authentique semble favoriser le développement des compétences dans un contexte de formation entièrement composée de stages au niveau des Études supérieures. En effet, de façon générale, les ÉFPC semblent soutenir plusieurs hypothèses que l'évaluation authentique préconise. En premier lieu, une méthode d'évaluation basée sur des performances complexes, lorsqu'elle est utilisée de façon répétitive et sur une courte période de temps, semble permettre de recueillir des données fiables pour inférer sur la compétence d'un individu. Dans notre étude, les commentaires des étudiants témoignent de cet élément. Ils reconnaissent que lorsque différents professeurs font des mêmes remarques sur les mêmes aspects de leur performances, même dans différentes situations d'évaluation, ils finissent par les croire et les accepter.

Deuxièmement, une évaluation avec des situations proches à des futures situations professionnelles de l'étudiant semble permettre une meilleure inférence sur les capacités de l'étudiant à résoudre d'autres problèmes dans les mêmes domaines cliniques, tout au moins selon l'opinion des professeurs. À ce propos, dans une perspective d'évaluation formative, l'utilisation des situations cliniques réelles représente un contexte d'évaluation valide aux yeux des professeurs et étudiants pour mesurer les performances de ces derniers. Aucun des participants n'ont contesté les situations d'évaluation utilisées.

En troisième lieu, l'utilisation d'une grille d'évaluation descriptive globale où sont inscrits de façon générique les tâches professionnelles et les niveaux de performance désirables et indésirables facilite la communication des informations, à la fois aux

professeurs et aux étudiants. Aucun de nos professeurs n'a eu une formation préalable à la méthode d'ÉFPC et pourtant, une consistance interne au-dessus de 0.80 a été atteinte après trois mois d'utilisation.

Finalement, notre recherche confirme le rôle essentiel de la rétroaction dans l'évaluation formative, toujours dans une perspective de développement des compétences. Nos données ouvrent une voie dans la recherche plus poussée sur les stratégies pour développer les capacités à l'auto-évaluation. Nos données sur la co-évaluation semblent alors démontrer que c'est en cherchant à comprendre la perception de l'étudiant de ses propres performances afin de cibler la rétroaction que l'on réussirait à mieux l'aider à développer son auto-évaluation; une étape essentielle dans le développement de la métacognition et l'atteinte de l'autonomie professionnelle.

Mais, les limites de notre recherche ne permettent pas encore de vérifier le nombre de situations d'évaluation requises pour certifier que l'étudiant a une compétence professionnelle suffisante pour entreprendre une carrière professionnelle indépendante. Il reste aussi à vérifier la faisabilité de cette méthode d'évaluation lorsqu'elle est utilisée sur une plus grande échelle avec des professeurs et des étudiants dans un contexte régulier de formation et qui n'ont pas de biais favorable comme celui que les participants à notre recherche pourraient avoir. Il serait tout aussi intéressant de vérifier dans une étape ultérieure si l'utilisation à large échelle d'une grille d'évaluation descriptive globale telle que celle des ÉFPC réussit à maintenir un niveau élevé de fidélité de mesure même s'il n'y a pas de formation préalable des professeurs. Ceci répondrait aux besoins d'une méthode d'évaluation fiable et valide dans un contexte de formation par stages où un nombre élevé de professeurs est impliqué et où la variabilité des problèmes cliniques restent une contrainte réelle.

SOURCES DOCUMENTAIRES

ACGME (1999). ACGME Outcome Project, General competencies, <http://www.acgme.org/outcome/comp/compfull.asp>, accédé le 28 février 2006.

Allal, L. (1981). Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application. In Allal, L., Cardinet J., Perrenoud P. (dir.), *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.

Allal, L. (1991). *Vers une pratique de l'évaluation formative; Matériel de formation continue des professeurs*. Belgique : De Boeck & Larcier.

Allal, L. (1993). Régulation métacognitive : quelle place pour l'élève dans l'évaluation formative ? In Allal, L., Bain, D et Perrenoud, P. (dir.), *Evaluation formative et didactique du français*. Neuchâtel : Delachaux et Niestlé.

Allal, L. (1999). Impliquer l'apprenant dans le processus d'évaluation; promesses et pièges de l'auto-évaluation. In C. Depover et B. Noël (dir.), *L'évaluation des compétences et des processus cognitifs*, pp 35-56. Bruxelles : De Boeck.

Albrecht, R. (1991). *L'évaluation formative, une analyse critique*. Bruxelles : De Boeck.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington D.C. : American Educational Research Association.

Andrich, D. (1997a). Rating scale analysis. In J. P. Keeves (dir.), *Educational research, methodology and measurement : an international handbook*. (2^e ed.), 874-880. Australie : Pergamon.

Andrich, D. (1997b). Thurstone scale. In J .P. Keeves (ed.) : *Educational research, methodology and measurement :an international handbook*, (2^e ed.), 819-821. Australie : Pergamon.

Angoff, W. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (dir.) *Educational measurement*, (2^e ed.). Washington DC : American Council of Education.

Arnold, L., Willoughby T. L., Calkins E. V. (1985). Self-evaluation in undergraduate medical education : a longitudinal perspective. *Journal of Medical Education*, 100 (1) : 21-28.

Battistone, M. J., Mine C., Sande M. A., Pangaro L. N., Hemmer P. A., Shomaker T. S. (2002). The feasibility and acceptability of implementing formal evaluation sessions and using descriptive vocabulary to assess student performance on a clinical clerkship. *Teaching & Learning in Medicine*, 14 (1) : 5-10 .

Beaumier, A., Bordage G., Saucier D., Turgeon J. (1992). The nature of the clinical difficulties of first year family medicine residents under direct observation. *Canadian Medical Association Journal*, 146(4) : 489-495.

Bejar, I. (1983). Subject matter experts' assessment of item statistics. *Applied Psychological Measurement*, 7 : 303-310.

Ben-David, M. F. (1999). AMEE Guide No 14 : Outcome-based education: part 3- Assessment in outcome-based education. *Medical teacher*, 21(1):23-25.

Biernat, K., Simpson D., Duthie E., Dawn B., London R. (2003). Primary care residents self-assessment skills in dementia. *Advances in Health Sciences Education*, 8 : 105-110.

Black, P., William D. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7-74.

Blackstein, D., Agnidis Z., Regehr R., Reznick, R. (2004) . The effectiveness of video feedback in the acquisition of orthopedic technical skills. *American Journal of Surgery*, 187 : 427-432.

Bloom, B. S. & coll. (1956). *Taxonomy of educational objectives : the classification of educational goal*. New-York : Mc Kay [cité dans de Landsheere V., 1984].

Bloom, B. S., Hasting J. T., Madaus G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New-York : Mc Graw Hill.

Bordage, G. (1999) . Why did I miss the diagnosis ? Some cognitive explanations and educational implications. *Academic Medicine*, 74 (10) : S138- S143.

Borman, E., Grady P. O. (1997). Postgraduate training policy paper prepared by the subcommittee on post graduate training; Permanent working group of European Junior Hospital Doctors . *Medical Education.*, 341: 3-8.

Boud, D. (2000). Sustainable assessment : rethinking assessment for the learning society. *Studies in Continuing Education*, 22 (2) : 151-167.

Boulet, J. R.; McKinley, D. W.; Norcini, J. J.; Whelan, G. P. (2002). Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Advances in Health Sciences Education*, 7 : 85-97.

Brien, R. (1981). *Design pédagogique : Introduction à l'approche de Gagné et de Brigg*. Québec : les éditions St-Yves.

Brinko, K. T. (1993). The practice of giving feedback to improve teaching : what is effective ? *Journal of Higher Education*, 64 (5) : 574-593.

Brooks, L. R., Norman G. R., Allen S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology : General*, (120) : 278-287.

- Butler, R. (1988). Enhancing and undermining intrinsic motivation : the effects of task-involving and ego- involving evaluation on interest and performance. *British Journal of Educational Psychology* , 58 :1-14.
- Calhoun, J. G., Ten Haken J. D .& Woolliscroft J. O. (1990). Medical students development of self- and peer-assessment skills : a longitudinal skill. *Teaching & Learning in medicine*, 2 (1) :25-29.
- Carline, J. D., Paun D. S., Thiede K. W., Ramsey P. G. (1992). Factors affecting the reliability of ratings of students' clinical skills in medicine clerkship. *Journal of General Internal Medicine*, 7 : 506-510.
- Chi, M. T. H., Glaser R., Rees E. (1982). Expertise in problem solving. In R. J. Sternberg (dir.), *Advances in the psychology of human intelligence*, 161-183. Hillsdale, N. J. : Lawrence Erlbaum Associates. [cité par Tardif , 1997].
- Cogbill, K. K., O'Sullivan P. S., Clardy J. (2005). Residents' perception of effectiveness of twelve evaluation methods for measuring competency. *Academic Psychiatry*, 29 (1) : 76-81.
- Cooksey, J., York J. (1999). Lessons learned from the teaching hospitals and the medical education model. *Peabody Journal of Education*, 74 (3 & 4) : 21-32.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. California : Wadsworth.
- Cushing, A. (2002). Assessment of non-cognitive factors. In G. R. Norman, C. P. M. Vander Vleuten, D. I. Newble (dir.), *International handbook of research in Medical Education*, 711-756. Boston : Kluwer Academic Publishers.

Dauphinee, W. D., Blackmore D. E., Smee D., Rothman A. I. & Reznick R. (1997). Using the judgments of physician examiners in setting the standards for a national multi-center high stakes OSCE. *Advances In Health Sciences Education*, (2) : 201-211.

Day, S. C., Grosso L. J., Norcini J. J., Swanson D. B., Horne M. H. (1990). Residents' perception of evaluation procedures used by their training program. *Journal of General Internal Medicine*, 5 (5) : 421-426.

De Landsheere V. & De Landsheere G. (1984). *Définir les objectifs de l'éducation*, (5^e ed.). Paris : Presses universitaires de France.

Des Marchais, J. (1996). *Apprendre à devenir médecin, bilan d'un changement pédagogique centré sur l'étudiant*. Canada : Presse de l'Université Sherbrooke.

Downing , S.M. (2002). Assessment of knowledge with written test forms. In G. R. Norman, C. P. Van Der Vleuten, D. J. Newble, *International handbook of research in medical education*. Great Britain : Kluwer Academic Publishers.

Dunn-Rankin, P. & Zhang S. (1997). Scaling methods. In J. P. Keeves (dir.), *Educational research, methodology and measurement : an international handbook*, (2^e ed.), 790-798. Australie : Pergamon.

Durning, S. J., Cation L. J., Markert R. J., Pangaro L. N. (2002). Assessing the reliability and validity of the mini-clinical evaluation exercise for Internal Medicine residency training. *Academic Medicine*, 77 :900-904.

Earl, L. (2003). *Assessment as learning : using classroom assessment to maximize student learning* : Thousand Oaks, Corwin Press.

Ebel, R. L. (1972). *Essentials of Educational Measurement*. New-Jersey : Prentice-Hall, Englewood Cliffs.

Embretson, S. (1983) . Construct validity; Construct representation versus nomothetic span. *Psychological Bulletin*, 93 : 179-197.

Ende, J. (1983). Feedback in clinical medical education. *JAMA*, 250 : 777-781.

Erwin, T. D. & Wise S. (2001). Standard setting. In R.A.Voorhees (dir.), *Measuring what matters: competency-based learning models in Higher Education. New directions for institutional research*, (110) : 55-64.

Ewell, P. T. (1985). Assessing Educational Outcomes. *New directions for Institutional Research*, 47. San-Francisco : Jossey-Bass.

Falchikov, N.; Boud, D. (1989). Students self-assessment in Higher Education : a meta analysis. *Review of Educational Research*, 59 (4) : 395-430.

Fallows, S., Chandramohan B. (2001). Multiple approaches to assessment : reflections on use of tutor, peer, and self-assessment. *Teaching in Higher Education*, 6 (2) :229-246.

Farrell, S. E. (2005). Evaluation of student performance : Clinical and professional performance. *Academic Emergency Medicine*, 12 (4) : 302.c6-302.c10.

Ferland, J. J. (1987). *Les grandes questions de la pédagogie médicale, perspective nord-américaine*, chap. 1 à 5. Québec : Presses de l'Université Laval.

Fleming, N. D. (1999). Biases in marking of student's written work quality. In S. Brown & A. Glassner (dir.), *Assessment matters in Higher Education : choosing and using diverse approaches*. United Kingdom : SRHE & Open University Press.

Flexner, A (1910). *Medical Education in the United States and Canada : a report to the Canergie Foundation for the Advancement of teaching*, bull. 4. New-York : The Canergie Foundation.

Frank, K. J. R., Jabbour M., Tugwell P. & al.(1996). Skills for the new millenium : report of the societal needs working group; CanMEDS 2000 Project. *Annals of Royal College of physicians and surgeons of Canada*, 29 : 206-216.

Friedman, M. & Mennin S. P. (1991). Rethinking critical issues in performance assesment . *Academic Medicine*, 66 : 390-395.

Gagné, R. M., Briggs L. J. (1974). *Principles of instructional design*. New-York : Holt, Rinehart & Winston.

Gagné, R. M. (1985). *The conditions of learning*, (4th ed.). New-York : Holt, Rinehart & Winston.

Gagné, R. M., Wager W. W., Grolas K. C., Leller J. M. (2005). *Principles of instructional design*, (5th ed.). New-York : Thomson Wadsworth.

Gardner, H. (1992). Assessment in context : the alternative to standardized testing; In B. R. Gifford & M. C. O'Connor (dir.), *Changing assessments : alternative views of aptitude, achievement and intruction*, 77-121. Boston : Kluwer Academic Publishers.

Garibaldi, R. A., Suhayah R., Moore M., Waxman, H., (2002). The in-training examination of Internal Medicine an analysis of resident performance overtime. *Annals of Internal Medicine*, (137) : 505-510.

General Medical Council (1977). Objectives in Medical Education. *Medical Education*, (11) : 241-243.

Gipps, C. & Stobart G. (2003). Alternative assessement. In T. Kellaghan, D. L. Stufflebeam (dir.) *International handbook of educational evaluation*. Boston : Kluwer Academic Publishers.

Glasner, A. (1999). Innovations in student assessment : a system wide perspective. In S. Brown & A. Glasner (dir.), *Assessment matters in Higher Education : choosing and using diverse approaches*. USA : SRHE & Open University Press.

Gordon, M. J. (1991). A review of the validity and accuracy of self-assessment in health professions training. *Academic Medicine*, 66 (12) : 762-769.

Gordon, M. J. (1992). Self-assessment programs and their implications for health professions training. *Academic Medicine*, 67 (10) : 672- 679.

Gray, J. (1996). Global rating scales in residency education. *Academic Medicine*, 77 (Jan. Sup.), 5 : 55-63.

Hambleton, R. K. (1997). Standard setting in criterion – referenced tests. In P. Keeves (dir.) : *Educational research, methodology and measurement : an international handbook*, (2^e ed.), 798-803 . Australie : Pergamon.

Hampton, J. R., Harrison M. J. G., Mitchell J. R. A., Prichard J. S., Seymour C. (1975). Relative contributions of history – taking, physical examination and laboratory investigation to diagnosis and management of medical outpatients. *British Medical Journal*, 2 : 486-489.

Harden, R.M., Crosby, J. R, Davis, M. H. (1999). AMEE guide no 14 : Outcome-based education, part 1- An introduction to outcome-based education. *Medical Teacher*, 21 (1) : 7-14.

Harden, R. M. (2002). Learning outcomes & instructional objectives : is there a difference ? *Medical Teacher*, 24 (2) : 151-155.

Harlen, W & James M. (1997). Assessment and learning; Differences and relationships between formative and summative assessment. *Assessment in Education*, 4 (3) : 365-379.

Hattie, J. (1987). Identifying the salient factors of a model of student learning : a synthesis of meta-analyses. *International journal of Educational Research*, 11 : 187-212.

Hattie, J. (1998). Assessment and classroom learning : a deductive approach. *Assessment in Educatio*, 5 (1) : 111-122.

Herbert, W. P., Mc Gaghie W. C., Droegemueller W., Riddle M., Maxwell K. L. (1990). Student evaluation in obstetric and gynecology : self- versus departmental assessment. *Obstetrics & Gynecology*, 76 : 458-461.

Hodder, R. V., Rivington R. N., Calcutt L. E. & Hart I. R. (1989). The effectiveness of immediate feedback during the Objective Structured Clinical Examination; *Medical Education*, 23, 184-188.

Hodges, B., Regehr G., Mc Naughton N., Tiberius R., Hanson M. (1999) . OSCE checklists do not capture increasing levels of expertise . *Academic Medicine*, 74 : 1129-1134.

Holm, H. A. (2002). Post-graduate Education. In G. R. Norman, C.P.M. Van der Vleuten, D. I. Newble (dir.), *International handbook of research in medical education*, 381-413. Boston : Kluwer Academic Publisher.

Holmboe, E. S., Hawkins R. E., Hammett T. W., Mackkrell-Gaglione M. (1996). Current methods in the evaluation of clinical competence. *Society of General Internal Medicine annual meeting* (1996).

Holmboe, E. S., Hawkin, R. E. (1998). Methods for evaluating the clinical competence of residents in Internal medicine : a review. *Annals of Internal Medicine*, 129 : 42-48.

Holmboe, E. S.; Huot, S.; Chung, J.; Norcini, J.; Hawkins, R. E. (2003). Construct validity of the mini Clinical Evaluation Exercise (mini CEX). *Academic medicine*, 78 (8) : 826-830.

Holmboe, E. S. (2004). Faculty and the observation of trainees' clinical skills : problems and opportunities . *Academic Medicine*, 79 (1) : 16-22.

Huba, M. E. & Freed J. E. (2000). *Learner – centered assessment on college campuses, shifting the focus from teaching to learning*. USA : Allyn & Bacon.

Jolly, B. C. (2002). Faculty development for curricular implementation. In G. R. Norman, C. P. M. Van der Vleuten & D. I. Newble (dir.), *International handbook of research in Medical Education*, 945-968. Boston : Kluwer Academic Publishers.

Jouquan, J. (2002) : L'évaluation des apprentissages des étudiants en formation médicale initiale. *Pédagogie Médicale*, 3 (1) : 38-52.

Jouriles, N., Burdick N., Hobgood C. (2002). Clinical assessment in emergency medicine. *Academic Emergency Medicine*, 9 (11) : 1289-1294.

Kane, M. (1992). The assessment of professional competence. *Evaluation & The Health Professions*, 15 (2) : 163-182.

Kane, M. (1999). Validating measures of performance. *Educational measurement; issues & practice*, 18(2) : 5-17.

Karle, H., Nystrup J. (1995). Comprehensive evaluation of specialist training : an alternative board examination in Europe. *Medical Education*, 29 : 38-46.

Khan, K. S., Davies D. A. & Gupta J. K. (2001). Formative self - assessment using multiple true- false questions on the Internet feedback according to confidence about correct knowledge. *Medical Teache* , 23 : 158-163.

Kirch, W, Schafil C. (1996). Misdiagnosis at a university hospital in 4 medical eras – report on 400 cases . *Medicine*, 75 (11) : 29-40.

Klenowski, V. (1995). Student self-evaluation processes in student – centred teaching and learning contexts of Australia and England. *Assessment in Education*, 2 (2) : 145-163.

Klessig, J. M., Wolfsthal. S. D. & Levine M. A., Stickley W., Bing-You R., Lansdale T. F., Battinelli D. L. (2000) . A pilot survey study to define quality in residency education. *Academic Medicine*, 75 (1) : 71-73.

Kolm, P., Verhurst, S. J. (1984). Comparing self and supervisor evaluations : a different view. *Proceedings of Annual Conference on Research in Medical Education*, 23 : 63-67.

Laurier, M. D., Toussignant R., Morissette D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages*, (3^e ed.). Québec : Gaetan Morin.

Laveault, D., Leblanc R. & Leroux, J. (1999). Autorégulation de l'apprentissage scolaire : interaction entre processus métacognitifs et déterminants de la motivation. In C. Depover & B. Noël (dir.), *L'évaluation des compétences et des processus cognitifs. Modèle, pratiques et contextes*, 81-98. Bruxelles : De Boeck.

Leach, D. C. (2001) . Changing education to improve patient care. *Quality in Health car*, 10, (suppl. II) : ii 54-ii 58.

Le Boterf, G. (1995). *De la compétence:essai sur un attracteur étranger*. Paris : Les Editions d'Organisation.

Le Boterf, G. (2000). *Construire les compétences individuelles et collectives*. Paris : Les Editions d'Organisation.

Legendre, R. (1993). *Dictionnaire actuel de l'éducation*, (2^e ed.). Montréal : Guérin & Paris, ESKA.

Lewis, D. M., Green D. R., Mitzel H. C., Baum K. & Patz R. J. (1998) The Bookmark standards setting procedure : methodology and recent implementations. Cité par Erwin T. D. & Wise S. L. (2001) Standard setting. In R Voorhees (dir.) : *Measuring what matters : competency-based learning models in higher education : New Direction For Institutional research*, 110 : 55-64.

Lin, J. (2005). The bookmark standard setting procedure; strength and weaknesses; Site web:

http://www.educational.valberta.ca/educ/psych/crame/files/standard_setting.pdf,
consulté : février 2006

Linn, R. L. (2000). Assessment and accountability. *Educational Researcher*, 29 (2) : 4-15.

Louis, R., Hensler H. (2003). L'évaluation des compétences professionnelles à l'enseignement : perspective nouvelle et implication sur les plans théorique et pratique. In Laurier, M. (dir.), *Évaluation & communication, de l'évaluation formative à l'évaluation informative*, 191-235. Québec : Les éditions Quebecor.

Mager, R. F. (1975) . *Preparing objectives for instruction*. Belmont CA : Fearon.

Mc Avoy (1985). How to choose and use educational objectives . *Medical Teacher*, 7 (1) : 27-35.

Mc Gaghie, W., Miller G., Sajid A., Telder T. (1978). Introduction à un enseignement médical fondé sur l'acquisition des compétences. *Cahier de santé publique*, no. 68, Organisation mondiale de la santé.

Mattheos, N., Nattestad A., Falf-Nilson E., Attström R. (2004). The interactive examination : assessing students' self-assessment ability . *Medical Education*, 38 : 378-389.

Maudsley, R. F. (1989). Effective in-training evaluation. *Medical Teacher*, 11 (3/4) : 285-290.

Mavis, B. E.; Henry R. C. (2002). Between a rock and a hard place : finding a place for the OSCE in medical Education . *Medical Education*, 36 : 408-409.

Messick, S. (1993a). Foundations of validity : meaning and consequences in psychological assessment. Research report. *Educational Testing Service*. New Jersey : Princeton.

Messick, S. (1993b) . Validity. In R. L. Linn, *Educational measurement*, (3^e ed.), 13-104. New-York : American Council on Education & Oryx Press.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments . *Educational Researcher*, 23 (2) : 13-23.

Miller, G. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65 (9) : 563-567.

Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28 (4) : 383-394.

Mislevy, R. J., Wilson M. R., Ercican K., Chudowsky N. (2003). Psychometric principles in student assessment. In T. Kellagham, D. Stufflebeam, L. A. Wingate. (dir.), *International handbook of Educational Evaluation*, 489-532. Dordrecht : Kluwer Academic Publishers.

Mitzel, H. C. , Lewis D. M., Patz R. J. & Green D. R. (2001). The bookmark procedure : psychological perspectives. In G. J. Cizek (dir.), *Setting Performance Standards : concept methods and perspectives*, 249-281. New Jersey : Erlbaum.

Morgenstern, C. & Keeves J. P. (1997). Descriptive scales. In J. P. Keeves (dir.) *Educationnal Research, Methodology and measurement : an international handbook* (2^e ed.), 900-908. Australie : Pergamon.

Nanson, E. M. (1975). Medical Education 100 years after. *New-Zealand Medical Journal*, (81) : 134-139.

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational Psychology Measurement*, 14 : 3-19 [cité par Hambleton, 1997].

Needham, D. R., Begg I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer : memory-oriented training promotes memory for training. *Memory and Cognition*, 19 (6) : 543-547.

Newell, A., Simon H. A. (1972). Human problem solving. Englewood Cliffs, N.-J. Prentice-Hall.

Nguyen, D. Q. (2006a). Approche par objectifs ou approche par compétences : quelles implications pour l'enseignement et pour l'évaluation dans la formation clinique ? Article qui sera soumis à la revue *Pédagogie médicale*.

Nguyen, D. Q. (2006b). Modélisation d'une grille d'évaluation formative des performances dans une perspective d'approche par compétences : un exemple de formation post-doctorale en Médecine Interne. Article qui sera soumis à la revue *Mesure & Évaluation en Éducation*.

Norcini, J. J., Shea J. A. & Kanya D. T. (1988). The effect of various factors on standard-setting. *Journal of Educational measurement*, 25 : 57-65.

Norcini, J. J. & Shea J. A. (1992). The reproductibility of standards over groups and occasions. *Applied Measurement in Education*, 5 : 63-72.

Norcini, J. J., Blank L. L., Aznold J. K., Kimball H. R. (1995). The mini-CEX (Clinical evaluation exercise) : a preliminary investigation. *Annals of Internal Medicine*, 123 : 795-799.

Norcini, J. J.; Blank, L. L.; Aznold, G. K.; Kimball, H. R. (1997). Examiner differences in the mini-CEX. In J. J. A. Scherpbier; C. P. M. Van der Vleuten; J. J. Rethans, A. F. W. Van der Steeg (eds) : *Advances in medical education*, 170-172: Dordrecht : Kluwer Academic Publishers.

Norcini, J. J. & Guille R. (2002). Combining tests and setting standards. In G. R. Norman, C. P. M. Vander Vleuten, D. I. Newble (dir.), *International handbook of research in Medical Education*, 811-834. Great Britain : Kluwer Academic Publishers.

Norcini, J. J., Blank L. L., Duffy S., Fortna G. (2003) : The mini-CEX : A method for assessing clinical skills. *Annals of Internal Medicine*, 138 : 476-481.

Norton, L. (2004). Using assessment criteria as learning criteria : a case study in psychology. *Assesement & Evaluation in Higher Education*, 29, 6 : 687-702.

Nunnally, J. C. (1978). *Psychometric theory*; chapitre 2 : Scaling models, (2^e ed.). New-York : Mc Graw-Hill.

Orsmond, P., Merry S., Reiling K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self - assessment. *Assessment and Evaluation in Higher Education*, 27 (4) : 309-323.

Ouellet, A. (2003). Pour une évaluation des apprentissages en lien avec la compétence. In M. D. Laurier (dir.), *Évaluation et communication : de l'évaluation formative à l'évaluation informative*, 59-99. Québec : Éditions Québecor.

Pangaro, L. (1999). A new vocabulary and other innovations for improving descriptive in-training evaluations. *Academic Medicine*, 74 (11) : 1203-1207.

Papadakis, M., Osborn E. H., Cooke M., Healy K. (1999). A strategy for detection and evaluation of unprofessional behavior in medical student. *Academic Medicine*, (74) : 980-990.

Pellegrino, J. W. (1992). Understanding what we measure and measuring what we understand. In B. R. Gifford & M. C. O'Connor (dir.), *Changing assessments : alternative views of aptitude, achievement and instruction*, 275-301. Boston : Kluwer Academic Publishers.

Peterson, M. C., Holbrook J. H., Hales D. V., Smith N. L., Staker L.V. (1992). Contributions of the history, physical examination and laboratory investigation in making diagnosis. *Western Journal of Medicine*, 156 : 163-165.

Perrenoud, P. (1981). Les différences culturelles aux inégalités scolaires : l'évaluation et la norme dans un enseignement indifférencié. In L. Allal, J. Cardinet., P. Perrenoud (dir.), *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.

Perrenoud, P. (1997). *Construire des compétences dès l'école*. Paris : ESF.

Petrusa, E. R. (2002). Clinical performance assessments. In G. R. Norman, C. P. M. Vander Vleuten, D. I. Newble (dir.), *International handbook of research in Medical Education*, 673-710. Great Britain : Kluwer Academic Publishers.

Popham, W. J. (2005). *Classroom assessment : What teachers need to know*, (4^e ed.). California : Pearson Education.

Pugh, E. W., Lloyd G. J., Mc Intyre N. (1975) . Relevance of educational objectives for medical education. *British Medical Journal*, (3) : 688-691.

Quirk, M., Stone S., Chuman A., Devaney-O'Neils S., Mazor K., Starr S., Lasser D. (2002). Using differences between perceptions of importance and competence to identify teaching needs of primary care preceptors. *Teaching and Learning In Medicine*, 14 (3) : 157-163.

Regehr, G., Norman G. R. (1996) . Issues in cognitive psychology : implications for professional education. *Academic Medicine*., 71(9) : 988-1001.

Regehr, G., MacRae H., Reznick R. K., Szalay D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination . *Academic Medicine*, 73 : 993-999.

Resnick, L.B. & Resnick D. L. (1992). Assessing the thinking curriculum : new tools for educational reform. In B. R. Gifford & M. C. O'Connor (dir.), *Changing assessments : alternative views of aptitude, achievement and instruction*, 37-77. Boston : Kluwer Academic Publishers.

Reznick, R. K., Blackmore D., Dauphinee W. D., Rothman A. I. & Smee S. (1996). Large scale high stakes testing with OSCE. Report from the *Medical Council of Canada. Academic Medicine*, 71(1) : 519-521.

Roegiers, X. (2001). *Une pédagogie de l'intégration : compétences et intégration des acquis dans l'enseignement*, (2^e ed.). Bruxelles : De Boeck.

Rushton, A. (2005). Formative assessment; a key to deep learning. *Medical Teacher*, 27 (6) : 509-513.

Saddler, R. (1998). Formative assessment : revisiting the territory. *Assessment in education*, 5 (1), 77-84.

Scallon, G. (1988). *L'évaluation formative des apprentissages, la réflexion (tome 1) et l'instrumentation (tome 2)*. Québec : Presses de l'Université de Laval.

Scallon, G. (2000). *L'évaluation formative*. Québec : Éditions. du Renouveau pédagogique.

Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Québec : Éditions du Renouveau pédagogique.

Scriven, M. (1967) . The methodology of evaluation. In Perspectives of curriculum evaluation, *AERE monographs series on Curriculum Evaluation*, 1 : 39-83. Chicago : Rand McNally.

Shea, J., Fortna G. S. (2002). Psychometric methods. In G. R. Norman, C. P. M. Van der Vleuten, D. I. Newble (dir.), *International Handbook of Research in Medical Education*, 97-126 . Boston : Kluwer Academic Publishers.

Shepard, L., Glaser R. & Bohrnsteoft G. (1993). *Setting performance standards for student achievement*. California : National Academy of Education.

Schmidt, H. G., Van der Molen H. T. (2001). Self-reported competency ratings of graduates of a problem-based medical curriculum . *Academic Medicine*, 76 : 466-468.

Schuwirth, L. W. T., Van der Vleuten, C. P. M. (2004). Different written assessment methods : what can be said about their strenghs and weaknesses ? *Medical Education*, (38) : 974-979.

Son, L. K., Schwartz B. L. (2002). The relatioonship between metacognitive monitoring and control. In T. J. Perfect & B. L. Schwartz (ed.), *Applied Metacognition*, 15-39. United Kingdom : Cambridge University Press.

Stevens, D. D. & Levi , A. J. (2005). Introduction to rubrics, an assessment tool to save grading time, convey effective feedback and promote student learning. Virginia : Stylus.

Swanson, D. B., Norman G. R., Linn R. L. (1995). Performance-based assessment : lessons from the Health Professions. *Educational Researcher*, 24 (5), 5-11, 35.

Tardif, J. (1993). L'évaluation dans le paradigme constructiviste. In R. Hivon (dir.), *L'évaluation des apprentissages : réflexions, nouvelles tendances et formation*. Sherbrooke : Editions de CRP.

Tardif, J. (1997). Pour un enseignement stratégique, l'apport de la psychologie cognitive. Québec : Les éditions Logiques.

Thorndike, R. M., Cunningham G. K., Thorndike R. L., Hagen E. F. (1991). *Measurement and evaluation in psychology and education*, (5th ed.). New-York : Mac Millan Publishing company.

Tillema, H., Kessels W., Merjers F. (2000). Competencies as a building blocks for intergrating assessment with instruction in vocational education : a case from the Netherlands. *Assessment & Evaluation in Higher Education*, 25 (3) : 265-278.

Toohy, S.; Ryan G.; & Hughes C. (1996). Assessing the praticum. *Assessment & Evaluation in Higher Education*, 21 (3) : 215-227.

Torrance, H., Pryor J. (2002). Investigating formative assessment : teaching, learning and assessment in the classroom. Maidenhead : Open University Press.

Turnbull, J., Gray, J., MacFadyen J. (1998). Improving in-training evaluaion programs. *General Internal Medicine*, 13 : 316-328.

Turnbull, J. , Gray J., MacFadyen J., Van Barneveld C., Norman G. (2000). Clinical work sampling : a new approach to the problem of in-training evaluation. *Journal of General Internal Medicine*, 15 : 556-561.

Turnbull, J., Van Barneveld C. (2002) : Assessment of clinical performance : in-training evaluation. In G. R. Norman, C. P. M. Van der Vleuten, D. I. Newble (dir.), *International Handbook of Research in Medical Education*, 793-810. Boston : Kluwer Academic Publishers.

Usatine, R. P, Edelstein R. A., Yajima A., Slavin S. J., Wilkes M. S. (1997). Medical student perceptions of the accuracy of various new clinical evaluation methods. In A. J. J. Scherpbier, C. P. M., Van der Vleuten J. J., Rethans A. F., Van der Steeg (dir.) *Advances in medical education*, 200-202. Boston : Kluwer Academic Publishers.

Upitis, R. (1999). Teacher education reform: putting experience first. *Teacher Education Quaterly*, 26 (2) : 11-19.

Villeneuve, L. (1994) . *L'encadrement du stage supervisé*. Québec : Éditions St-Martin.

Voorhees, R. A (2001). Competency based learning models : a necessary future. In R.A.Voorhees (dir.), *Measuring what matters: competency-based learning models in higher education. New directions for institutional research* (110) : 5-13. San-Francisco : Jossey-Bass.

Wakefield, J. (1985). Direct observation. In V. R. Newfield & G. R. Norman (dir.), *Assessing clinical competence; Springer Series on Medical Education*, vol. 7, 51-70.

Ward, M., Gruppen L. & Regehr G. (2002). Measuring self-assessment : current state of the art . *Advances in Health Sciences Education*, 7 : 63-80.

Ward, M., Mac Rae H., Schlachta C., Mamazza J., Poulin E., Reznick R., Regehr G. (2003). Resident self-assessment of operative performance. *American Journal of Surgery*, 185 : 521-524.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46 (7) : 41-47.

Wiggins, G. (1998). *Educative assessment : designing assessments to inform and improve student performance*. San-Francisco : Jossey-Bass.

Williams, R. G., Osbourne C. E. (1982). Medical teachers' perspectives on development and use of objectives. *Medical Education*, (16) : 68-71.

Wolf, R. M. (1997). Rating scales. In J. P. Keeves (dir.), *Educational research, methodology and measurement : an international handbook*, (2^e ed.), 959-965. United Kingdom : Elsevier Science Ltd.

Woolliscroft, J. O. (2002). Medical student clinical education. In G. R. Norman, C. P. M. Van der Vleuten, D. I. Newble (dir.), *International handbook of research in Medical Education*, 365-380. Boston : Kluwer Academic Publishers.

Yorke, M. (1998). The management of assessment in Higher Education. *Assessment & Evaluation in Higher Education*, 23 (2) : 101-116.

DES LIVRES SUIVANTS ONT AUSSI ÉTÉ CONSULTÉS :

Harvey , G. (2003). Excel 2003 pour les nuls , 9 livres en 1, traduit par R. Dumont & B. Jolivald , New-York : Wiley Publishing inc.

Howell, D.C.(1998). Méthodes statistiques en sciences humaines, traduit par M. Rogier. Bruxelles : De Boeck Université.

Kinnear, P., Gray, C. (2005). SPSS facile appliqué à la psychologie et aux sciences sociales, traduit par N. Huet. Bruxelles : De Boeck & Larcier.

Norušis , M. (2005). SPSS 13.0 , Guide to data analysis. New Jersey : Prentice Hall.

Rosner, B. (2000). Fundamentals of biostatistics, (5th ed.). California : Duxbury - Thompson Learning.

Van Der Maren, J.M. (1996). Méthodes de recherche pour l'éducation, (2^e ed.). Bruxelles : De Boeck & Larcier.

Van Der Maren, J.M (1999). La recherche appliquée en pédagogie , des modèles pour l'enseignement- méthodes en sciences humaines, Bruxelles, De Boeck & Larcier.

Annexes

Annexe I

Évaluation formative des performances cliniques

Documents remis aux professeurs

1. Description de la méthode des évaluations formatives des performances cliniques (ÉFPC)

Les ÉFPC sont des sessions d'évaluation des performances cliniques des résidents au cours de leur formation. Cette méthode se veut une méthode d'évaluation formative par une approche d'évaluation authentique des compétences cliniques dans un contexte de contraintes de temps et de disponibilité des professeurs pour des sessions de formation professorale préalable à l'activité.

Pour cela, chaque session de ÉFPC ne durera qu'une heure et portera sur le problème médical principal du patient (aucune revue globale des autres problèmes mineurs du patient ne sera bordée). Un accent particulier sera porté sur l'entrevue médicale où le résident doit voir un patient qui ne lui est pas connu préalablement et le résident devra performer comme un médecin spécialiste à qui on adresse un patient pour une première fois, avec un problème spécifique. Il devra alors questionner et examiner le patient et à partir de l'intégration de toutes ses connaissances scientifiques acquises antérieurement et des données cliniques rapportées par le patient il doit pouvoir résoudre le problème clinique qui lui est soumis. Afin de rendre la situation d'évaluation authentique, le patient sera un véritable patient, le problème sera un problème clinique véritable qu'un médecin peut rencontrer au cours de sa pratique professionnelle. Comme il s'agit d'une évaluation des compétences cliniques c'est-à-dire qu'en plus des domaines cognitifs, des domaines non-cognitives sont aussi vérifiés; celles relatives aux habiletés communicationnelles (i.e. : différentes techniques pour questionner lors d'une entrevue médicale ...) et aux habiletés psychomotrices (ex. : les manœuvres à faire pour examiner un patient ...). De plus, afin de respecter l'authenticité de la situation d'une consultation médicale, l'entrevue médicale se termine par des explications que le résident aura à donner au patient, immédiatement après le questionnaire et l'examen physique.

L'ÉFPC est une méthode d'évaluation formative des compétences cliniques, de telle façon que le cheminement de raisonnement clinique du résident pour solutionner le problème est important; pour cela, dans un deuxième temps, après la rencontre médicale, le résident doit résumer ce qu'il a retenu de toute la rencontre avec le patient, doit par la suite proposer un diagnostic et un diagnostic différentiel ainsi qu'un plan d'investigation et un plan thérapeutique en tentant d'intégrer des données scientifiques connues aux données cliniques rapportées. Cette étape est cruciale dans les ÉFPC, puisque cela permet au professeur de mieux saisir tout le processus mental de son étudiant et ses démarches cognitives de raisonnement, il verrait alors si le résident a su intégrer les symptômes rapportés par le patient, les signes cliniques, les impacts psychologiques de la maladie pour en arriver à un plan thérapeutique logique. De plus, des dimensions non-cognitives seront évaluées tout au long de la ÉFPC : la communication avec le patient, la communication avec ses pairs, l'organisation des réflexions, l'aspect empathique avec le patient lors de la proposition des traitements et des investigations. Cette compréhension plus approfondie permettra de rendre la rétroaction plus riche et plus bénéfique.

1.1 Déroutement d'une ÉFPC

Avant la session, le résident est rencontré par le professeur qui le met au courant du déroulement d'une ÉFPC et lui résume les attentes. Le résident reçoit alors préalablement une grille dont on se sert pour rubrique à l'évaluation, sur laquelle se trouvent les standards de performances, les niveaux de performances de façon descriptive et les dimensions qui seront évaluées. Au début de la session, le professeur présente le patient au résident et lui résume rapidement la raison pour laquelle le patient lui est adressé. Par la suite débute l'entrevue médicale sous l'observation directe.

Lors de l'entrevue, le professeur observe les performances cliniques du résident et évite d'intervenir au cours de l'entrevue médicale. L'observation directe à côté du patient et du résident est une méthode favorisée pour obtenir toutes les

informations sur la performance du résident avec le patient qui lui est présenté. Au cours de cette partie d'ÉFPC, le professeur utilise la grille comme guide pour cibler son attention sur les aspects à observer et il est aussi invité à annoter tous les incidents qui serviront d'incident critique pour des discussions de rétroaction plus tard.

Afin de rendre la méthode d'observation directe adéquate pour recueillir des informations concernant la performance des résidents, certains éléments concernant le processus sont nécessaires afin de s'assurer que la présence de l'observateur altère le moins possible la performance de l'étudiant : (Wakefeld, 1985); surtout lorsque l'observateur se retrouve dans le même local que le résident et le patient.

- ✓ L'observateur se tient en retrait et évite d'être dans la même ligne de vision à la fois du patient et du résident
- ✓ L'observateur n'intervient pas au cours de l'entrevue et évite à la fois de donner des conseils à l'étudiant ou de répondre au patient, afin d'éviter d'altérer la relation médecin-patient établie par le résident. Les seules interventions se réservent lorsque l'observateur juge que certaines manœuvres faites par le résident peuvent mettre en danger le patient (physique ou psychologique)
- ✓ La conclusion de l'observation directe devrait être une réassurance du patient, en l'absence du résident, afin d'éviter de biaiser le résident si possible, surtout lorsque l'évaluateur, qui connaît bien le patient se rend compte que l'explication donnée par le résident au patient est erronée (ex. : à un patient qui n'a qu'une migraine, à qui le résident dit qu'il pense qu'il a une tumeur, il est nécessaire alors de réassurer le patient).

Le résident est invité à se comporter autant que possible comme s'il n'était pas en situation d'évaluation; l'accent étant mis sur le fait que l'activité sert à l'enseignement, à l'aider à s'améliorer et non à une décision de réussite ou d'échec de stage ...

Le résident questionne le patient sur son problème principal, puis examine le patient en ciblant sur les systèmes rapportés par le patient comme ayant un problème (ex. : si le patient se plaint d'une douleur abdominale, l'examen physique portera principalement sur l'examen abdominal ...). Le but visé étant qu'à la fin, le résident doit pouvoir donner son opinion quant à l'explication de la douleur abdominale en question et ce qu'il fait pour aider le patient à s'en sortir. À la fin du questionnaire médical et de l'examen physique, le résident doit immédiatement soit répondre aux questions du patient, soit lui donner des explications. Ceci permet déjà de voir ce que le résident a compris des problèmes rapportés par le patient sans être biaisé par des opinions du professeur, en même temps de vérifier ses capacités de communiquer des informations médicales au patient.

Le patient, dans ces ÉFPC, est un patient connu médicalement par le professeur, et ce point est très important. Le professeur qui connaît bien son patient saura si son résident a réussi à trouver toutes les informations utiles et nécessaires, s'il a trouvé toutes les manifestations cliniques qu'il devrait trouver; et cela devient pertinent lors de la rétroaction. Il pourra ainsi déterminer si le résident a eu de la difficulté à résoudre le problème du patient dû à des problèmes de techniques d'examen physique ou d'entrevue et par la suite apprécier aussi la qualité du raisonnement clinique du résident, ou tout simplement par un manque de connaissances médicales préalables, les activités suggérées pour améliorer seront différentes.

Le professeur responsable de chaque ÉFPC est un médecin considéré spécialiste dans le domaine du problème du patient (i.e. un rhumatologue pour un problème clinique du système locomoteur tel que polyarthrite ...) et l'on évite de demander un médecin qui ne se considère pas comme compétent dans un domaine d'agir comme évaluateur d'une ÉFPC (ex. : un cardiologue évitera d'évaluer un résident au sujet d'un patient souffrant de douleur abdominale ...). Cet élément a été choisi expressément pour la qualité de la rétroaction; le professeur doit apprécier la qualité de la performance du résident, et, selon les écrits sur l'appréciation de la compétence; celui agissant comme évaluateur devrait être lui-même professionnel compétent dans le même domaine (ex. : l'on s'attend à ce qu'un ingénieur reconnaisse la compétence d'un autre ingénieur et l'on ne voit pas comment un médecin peut juger de la compétence d'un ingénieur). Cet élément n'a jamais été rapporté dans les écrits dans le domaine de la santé, ni sujet de recherche expressément porté. Mais, comme Brinko (1993) l'a mentionné, le professeur qui donne une rétroaction doit être crédible aux yeux de l'étudiant. La balise de temps établie pour le déroulement d'une ÉFPC est :

- 35 minutes de questionnaire et d'examen physique
- 5 à 10 minutes pour répondre aux questions du patient
- 5 minutes pour présenter le cas clinique de façon synthétique
- 5 minutes où le résident et le professeur remplissent la grille d'évaluation de façon indépendante
- 20 minutes de rétroaction.

Mais le professeur est évidemment responsable d'assouplir cette échelle de temps selon le besoin du résident.

Rétroaction

- Le résident et le professeur remplissent leur grille de façon indépendante après que le résident ait donné son rapport du cas clinique, sa proposition de diagnostic, son plan d'investigation ainsi que son plan thérapeutique. Par la suite, le résident et le professeur comparent leurs grilles et voient s'il y a des différences d'opinion et d'appréciation. La discussion qui s'ensuit vise à :
 - renforcer positivement le résident sur ce qu'il fait de façon adéquate pour qu'il maintienne l'acquis
 - améliorer ce qui est perçu comme « non-atteint » afin de développer des habiletés dans un but d'éventuellement développer des compétences.
 - Pour cela, une discussion respectant des principes de base du feedback aura lieu entre le professeur et le résident. Ces principes sont :
 - a) concernant le moment : immédiatement après la performance, lorsque cette performance est encore fraîche en mémoire (puisque aucun enregistrement n'a été fait)
 - b) Concernant la raison de la rétroaction : une insistance sur le but de l'activité clarifiera par la suite les informations et les commentaires transmis; le feedback sert à améliorer et à aider et non à sanctionner
 - c) Concernant le déroulement même (« comment » devrait avoir la forme l'activité ?) :
 - ✓ Sous forme de discussion
 - ✓ Dans un local où il n'y a pas de distraction

Dans un climat où le résident se sent à l'aise pour solliciter une rétroaction

- ✓ En tenant compte du besoin d'apprentissage du résident et non le besoin du professeur d'enseigner
- d) Concernant le contenu de la rétroaction
 - ✓ Les informations discutées seront les comportements observés, discutés sous l'angle des standards explicités sur la grille d'évaluation
 - ✓ Un partage d'opinion et une discussion ouverte où le professeur demande des questions qui permettent de mieux comprendre la raison de l'auto-appréciation du résident avant de discuter le point de vue du professeur
 - ✓ Cibler des informations spécifiques et éviter des perceptions générales vagues et imprécises (ex. : « le résident n'a pas laissé le patient finir les phrases et a ainsi eu de la difficulté à obtenir toutes les informations nécessaires » et non « vous êtes pressé et vous avez manqué toutes les informations utiles pour comprendre sa douleur »)
 - ✓ Cibler les performances et les comportements et éviter de porter le focus sur l'étudiant (ex. : dire « l'aspect des impacts financiers n'a pas été suffisamment approfondi pour comprendre pourquoi le patient n'a pas continué à prendre ses médicaments » au lieu de dire « vous semblez être mal à l'aise de poser des questions sur les aspects financiers et passez ainsi à côté la raison du manque de compliance de votre patient au traitement »)
 - ✓ Discuter sur ce qui est modifiable, sur ce que le résident peut contrôler et non sur ce qu'il ne peut pas changer : (ex. : « le patient parle souvent en monosyllable, vous pourriez alors lui poser des questions plus spécifiques telles que comment, quand ... sur les circonstances de la maladie pour obtenir des informations » plutôt que « vous devriez insister pour que le patient parle un peu plus sur les circonstances de sa maladie même s'il ne répond qu'en monosyllable »)

- ✓ Sélectionner des messages à passer et toujours penser à encadrer avec des éléments positifs un ou deux éléments négatifs; ce qui permet de mieux recevoir le message négatif et mieux le retenir. La surcharge d'information diminue et dilue l'importance de chaque message.
- ✓ Afin d'améliorer l'auto-évaluation, les éléments suivants doivent être sujets à une attention particulière lors de la discussion
- ❖ faire le lien entre les standards et l'intégration des connaissances, des savoir-faire et savoir-être pour déterminer comment l'on pourrait améliorer les interventions : ex. : si le résident n'a pas su utiliser des questions ouvertes pour laisser l'occasion au patient d'élaborer sur la chronologie de sa maladie de façon complète, ni inviter le patient à faire part de son vécu psychologique lors de l'apparition des symptômes; et que le résident, en s'auto-évaluant, a vu qu'il n'a pas tous les détails nécessaires pour arriver à un diagnostic plausible, une question possible du professeur serait : « selon toi, comment peut-on mieux obtenir des informations sur ce qui se passe avec ce patient depuis les six derniers mois ? » (le but de la question étant de faire réaliser au résident ce qu'il pourrait faire différemment)
- ❖ amener le résident à réfléchir sur ce qu'il a fait d'adéquat, et donc à maintenir, et l'aider à identifier les aspects à améliorer avec un plan de développement futur (insister pour qu'il propose des activités qui lui permettront de s'améliorer avant de lui proposer d'autres activités)
- La combinaison de l'auto-évaluation et de la rétroaction doit avoir comme but de développer la métacognition chez le résident : le rendre conscient de son processus cognitif, de ses forces et ses faiblesses et comment utiliser les ressources à sa disposition pour s'améliorer.

- La conclusion d'une ÉFPC devrait être une conclusion sur des activités pour améliorer les performances (ex : lecture additionnelle, discussion répétée avec d'autres professeurs, un stage électif clinique...) ainsi qu'un échéancier de réalisation. Le professeur peut remettre une copie de sa grille au résident .
- Par ailleurs, afin de s'assurer du caractère formatif de cet exercice, le professeur demandera au résident de lui faire part de la démarche entreprise pour s'améliorer entre les deux ÉFPC.

Annexe II

ÉVALUATION FORMATIVE DES PERFORMANCES CLINIQUES

Documents remis aux résidents

1. Description de la méthode des évaluations formatives des performances cliniques (ÉFPC)

Les ÉFPC sont des sessions d'évaluation des performances cliniques des résidents au cours de leur formation. Cette méthode se veut une méthode d'évaluation formative par une approche d'évaluation authentique des compétences cliniques dans un contexte de contraintes de temps et de disponibilité des professeurs pour des sessions de formation professorale préalable à l'activité.

Pour cela, chaque session de ÉFPC ne durera qu'une heure et portera sur le problème médical principal du patient (aucune revue globale des autres problèmes mineurs du patient ne sera bordée). Un accent particulier sera porté sur l'entrevue médicale où le résident doit voir un patient qui ne lui est pas connu préalablement et le résident devra performer comme un médecin spécialiste à qui on adresse un patient pour une première fois, avec un problème spécifique. Il devra alors questionner et examiner le patient et à partir de l'intégration de toutes ses connaissances scientifiques acquises antérieurement et des données cliniques rapportées par le patient il doit pouvoir résoudre le problème clinique qui lui est soumis. Afin de rendre la situation d'évaluation authentique, le patient sera un véritable patient, le problème sera un problème clinique véritable qu'un médecin peut rencontrer au cours de sa pratique professionnelle. Comme il s'agit d'une évaluation des compétences cliniques c'est-à-dire qu'en plus des domaines cognitifs, des domaines non-cognitives sont aussi vérifiés; celles relatives aux habiletés communicationnelles (i.e. : différentes techniques pour questionner lors d'une entrevue médicale ...) et aux habiletés psychomotrices (ex. : les manœuvres à faire pour examiner un patient ...). De plus, afin de respecter l'authenticité de la situation d'une consultation médicale, l'entrevue médicale se termine par des explications que le résident aura à donner au patient, immédiatement après le questionnaire et l'examen physique.

L'ÉFPC est une méthode d'évaluation formative des compétences cliniques, de telle façon que le cheminement de raisonnement clinique du résident pour solutionner le problème est important; pour cela, dans un deuxième temps, après la rencontre médicale, le résident doit résumer ce qu'il a retenu de toute la rencontre avec le patient, doit par la suite proposer un diagnostic et un diagnostic différentiel ainsi qu'un plan d'investigation et un plan thérapeutique en tentant d'intégrer des données scientifiques connues aux données cliniques rapportées. Cette étape est cruciale dans les ÉFPC, puisque cela permet au professeur de mieux saisir tout le processus mental de son étudiant et ses démarches cognitives de raisonnement, il verrait alors si le résident a su intégrer les symptômes rapportés par le patient, les signes cliniques, les impacts psychologiques de la maladie pour en arriver à un plan thérapeutique logique. De plus, des dimensions non-cognitives seront évaluées tout au long de l'ÉFPC : la communication avec le patient, la communication avec ses pairs, l'organisation des réflexions, l'aspect empathique avec le patient lors de la proposition des traitements et des investigations. Cette compréhension plus approfondie permettra de rendre la rétroaction plus riche et plus bénéfique.

1.1 Déroulement d'une ÉFPC

Avant la session, le résident est rencontré par le professeur qui le met au courant du déroulement d'une ÉFPC et lui résume les attentes. Le résident reçoit alors préalablement une grille dont on se sert pour rubriquer à l'évaluation, sur laquelle se trouvent les standards de performances, les niveaux de performances de façon descriptive et les dimensions qui seront évaluées. Au début de la session, le professeur présente le patient au résident et lui résume rapidement la raison pour laquelle le patient lui est adressé. Par la suite débute l'entrevue médicale sous l'observation directe.

Lors de l'entrevue, le professeur observe les performances cliniques du résident et évite d'intervenir au cours de l'entrevue médicale. L'observation directe à côté du patient et du

résident est une méthode favorisée pour obtenir toutes les informations sur la performance du résident avec le patient qui lui est présenté. Au cours de cette partie d'ÉFPC, le professeur utilise la grille comme guide pour cibler son attention sur les aspects à observer et il est aussi invité à annoter tous les incidents qui serviront d'incident critique pour des discussions de rétroaction plus tard.

Afin de rendre la méthode d'observation directe adéquate pour recueillir des informations concernant la performance des résidents, certains éléments concernant le processus sont nécessaires afin de s'assurer que la présence de l'observateur altère le moins possible la performance de l'étudiant : (Wakefeld, 1985); surtout lorsque l'observateur se retrouve dans le même local que le résident et le patient.

- ✓ L'observateur se tient en retrait et évite d'être dans la même ligne de vision à la fois du patient et du résident
- ✓ L'observateur n'intervient pas au cours de l'entrevue et évite à la fois de donner des conseils à l'étudiant ou de répondre au patient, afin d'éviter d'altérer la relation médecin-patient établie par le résident. Les seules interventions se réservent lorsque l'observateur juge que certaines manœuvres faites par le résident peuvent mettre en danger le patient (physique ou psychologique)
- ✓ La conclusion de l'observation directe devrait être une réassurance du patient, en l'absence du résident, afin d'éviter de biaiser le résident si possible, surtout lorsque l'évaluateur, qui connaît bien le patient se rend compte que l'explication donnée par le résident au patient est erronée (ex. : à un patient qui n'a qu'une migraine, à qui le résident dit qu'il pense qu'il a une tumeur, il est nécessaire alors de réassurer le patient).

Le résident est invité à se comporter autant que possible comme s'il n'était pas en situation d'évaluation; l'accent étant mis sur le fait que l'activité sert à l'enseignement, à l'aider à s'améliorer et non à une décision de réussite ou d'échec de stage ...

Le résident questionne le patient sur son problème principal, puis examine le patient en ciblant sur les systèmes rapportés par le patient comme ayant un problème (ex. : si le patient se plaint d'une douleur abdominale, l'examen physique portera principalement sur l'examen abdominal ...). Le but visé étant qu'à la fin, le résident doit pouvoir donner son opinion quant à l'explication de la douleur abdominale en question et ce qu'il fait pour aider le patient à s'en sortir. À la fin du questionnaire médical et de l'examen physique, le résident doit immédiatement soit répondre aux questions du patient, soit lui donner des explications. Ceci permet déjà de voir ce que le résident a compris des problèmes rapportés par le patient sans être biaisé par des opinions du professeur, en même temps de vérifier ses capacités de communiquer des informations médicales au patient.

Le patient, dans ces ÉFPC, est un patient connu médicalement par le professeur, et ce point est très important. Le professeur qui connaît bien son patient saura si son résident a réussi à trouver toutes les informations utiles et nécessaires, s'il a trouvé toutes les manifestations cliniques qu'il devrait trouver; et cela devient pertinent lors de la rétroaction. Il pourra ainsi déterminer si le résident a eu de la difficulté à résoudre le problème du patient dû à des problèmes de techniques d'examen physique ou d'entrevue et par la suite apprécier aussi la qualité du raisonnement clinique du résident, ou tout simplement par un manque de connaissances médicales préalables, les activités suggérées pour améliorer seront différentes.

Le professeur responsable de chaque ÉFPC est un médecin considéré spécialiste dans le domaine du problème du patient (i.e. un rhumatologue pour un problème clinique du système locomoteur tel que polyarthrite ...) et l'on évite de demander un médecin qui ne se considère pas comme compétent dans un domaine d'agir comme évaluateur d'une ÉFPC (ex. : un cardiologue évitera d'évaluer un résident au sujet d'un patient souffrant de douleur abdominale ...). Cet élément a été choisi expressément pour la qualité de la rétroaction; le professeur doit apprécier la qualité de la performance du résident, et, selon les écrits sur l'appréciation de la compétence; celui agissant comme évaluateur devrait être lui-même professionnel compétent dans le même domaine (ex. : l'on s'attend à ce qu'un ingénieur reconnaisse la compétence d'un autre ingénieur et l'on ne voit pas comment un médecin peut juger de la compétence d'un ingénieur). Cet élément n'a jamais été rapporté dans les écrits dans le domaine de la santé, ni sujet de recherche expressément porté. Mais, comme Brinko (1993) l'a mentionné, le professeur qui donne une rétroaction doit être crédible aux yeux de l'étudiant.

La balise de temps établie pour le déroulement d'une ÉFPC est :

- 35 minutes de questionnaire et d'examen physique
- 5 à 10 minutes pour répondre aux questions du patient
- 5 minutes pour présenter le cas clinique de façon synthétique
- 5 minutes où le résident et le professeur remplissent la grille d'évaluation de façon indépendante
- 20 minutes de rétroaction.

Mais le professeur est évidemment responsable d'assouplir cette échelle de temps selon le besoin du résident.

ANNEXE V GRILLE D'ÉFPC (4 niveaux)

Annexe V – Grille d'ÉFPC (4 niveaux)

Date : _____

Résident : _____

Grille remplie par : _____

Niveau : ☐ R-I ☐ R-II ☐ R-III

Recueillir des données cliniques avec le patient		Exemple, commentai et suggestions	
Inadéquat	Amélioration nécessaire	Adéquat	Excellent
1. Questionner le patient sur le problème médical principal	<p><input type="checkbox"/> Les détails cliniques sont obtenus de façon disparate, sans fil conducteur apparent</p> <p><input type="checkbox"/> Les informations sont obtenues de façon concises, preuve de connaissances physiopathologiques, mais, la chronologie des interventions antérieures n'a pas été obtenue</p>	<p><input type="checkbox"/> Les informations sont obtenues de façon concise, et la chronologie des interventions antérieures est obtenue de façon complète</p> <p><input type="checkbox"/> Toutes les informations ont été obtenues de façon concise avec un fil conducteur logique faisant preuve de connaissance approfondie de physiopathologie logiques sous-jacentes</p>	
2. Explorer des aspects psycho-sociaux de ce même problème médical	<p><input type="checkbox"/> Aucune vérification des impacts de la maladie sur la vie du patient</p> <p><input type="checkbox"/> Des habitudes de vie pertinentes ont été obtenues (ex. : tabagisme en relation avec la maladie cardiaque)</p> <p><input type="checkbox"/> Peu a été demandé sur les impacts de la maladie sur la vie du patient</p>	<p><input type="checkbox"/> Les habitudes de vie ainsi que les impacts de la maladie sur la vie du patient ont été adéquatement obtenus</p> <p><input type="checkbox"/> Toutes les informations sur l'impact de la maladie ainsi que la perception du patient vis-à-vis sa maladie ont été obtenues de façon adroite et respectueuse</p>	
3. Utiliser une technique d'entrevue appropriée pour obtenir des informations cliniques nécessaires	<p><input type="checkbox"/> Les questions trop dirigées empêchent des réponses complètes, et ne permettent pas l'établissement d'une relation de confiance</p> <p><input type="checkbox"/> Les différents types de questions (ouvertes, fermées, clarification, résumé ...) ont été utilisés pour obtenir des informations nécessaires, mais</p> <p><input type="checkbox"/> il a été peu tenu compte de l'état physique du patient (ex. : trop de questions à un patient faible)</p>	<p><input type="checkbox"/> Différents types de questions ont été adéquatement utilisés pour obtenir des informations nécessaires, en tenant compte de l'état clinique du patient, inspirant une relation de confiance</p> <p><input type="checkbox"/> En plus d'être adroit dans le questionnaire, a toujours été courtois et respectueux et a su établir une relation de confiance qui permet d'obtenir des informations complètes</p>	

4. Proposition d'un diagnostic différentiel		Exemples et commentaires		
	À améliorer	Bien	Très bien	
4.1	Établir un diagnostic différentiel avec une argumentation et une mise en priorité bien explicite.	Les solutions proposées sont logiques et la présentation du problème est fluide, se basant sur des données cliniques	Le diagnostic est logique, clair, élaboré avec une argumentation pertinente des données scientifiques tenant compte des données cliniques reçues	
5. Plan d'investigation				
	À améliorer	Bien	Très bien	
5.1	Proposer un plan d'investigation pur tenter de solutionner le problème clinique à l'évaluateur	Le plan est approprié aux diagnostics proposés et est rentable (a tenu compte des restrictions financières du patient et de l'établissement)	Le plan proposé a su intégrer des données scientifiques et cliniques (ce qui est idéal et ce qui est acceptable par le patient) et a aussi tenu compte des impacts sur le patient	
6. Proposition d'un plan thérapeutique				
	À améliorer	Bien	Très bien	
6.1	Établir un plan thérapeutique si jugé nécessaire à la fin de la rencontre	Les traitements proposés sont justifiés et faisant preuve de la réflexion entre le besoin de traitement immédiat et l'attitude expectative	Les traitements proposés sont argumentés en se basant sur des évidences scientifiques et tenant compte des données cliniques obtenues	
7. Communication d'information au patient				
	À améliorer	Bien	Très bien	
7.1	Répondre aux questions du patient	Les informations sont données de façon claire et détaillée	Les informations données sont claires, sans jargon avec un souci de s'assurer de la bonne compréhension du patient	

Durée de l'entrevue :	début :	fin :	Durée du feedback :
Commentaires généraux :	Points à maintenir		
Suggestions pédagogiques :			

Prochaine rencontre :	<input type="checkbox"/> nécessaire	<input type="checkbox"/> 2 semaines	<input type="checkbox"/> 4 semaines	<input type="checkbox"/> 6 semaines	<input type="checkbox"/> autres
-----------------------	-------------------------------------	-------------------------------------	-------------------------------------	-------------------------------------	---------------------------------

	Inadéquat	Amélioration nécessaire	Adéquat	Excellent
2. Proposer un diagnostic différentiel logique	<input type="checkbox"/> Le diagnostic proposé ne tient pas compte des problèmes cliniques rapportés par le patient, et <input type="checkbox"/> dont l'ordre des priorités n'est pas explicite	<input type="checkbox"/> Le diagnostic tient compte du problème clinique, faisant preuve de maîtrise des connaissances de base, mais <input type="checkbox"/> d'argumentation logique	<input type="checkbox"/> Le diagnostic proposé est logique avec une argumentation basée sur des données cliniques de façon juste et précise	<input type="checkbox"/> Le diagnostic est juste et précis, clair avec une argumentation intégrant à la fois les données cliniques et les connaissances scientifiques
3. Proposer un plan d'investigation pour tenter de solutionner le problème clinique du patient	<input type="checkbox"/> Le plan est inapproprié, n'ayant pas tenu compte du diagnostic proposé, et <input type="checkbox"/> pourrait mettre en danger le patient	<input type="checkbox"/> Le plan est approprié, mais <input type="checkbox"/> n'a pas tenu compte du patient (ses craintes, ses difficultés à se déplacer...), et <input type="checkbox"/> comporte des coûts potentiellement excessifs	<input type="checkbox"/> Le plan est approprié en tenant compte de l'impact sur le patient, et <input type="checkbox"/> des coûts au système de santé	<input type="checkbox"/> Le plan proposé a su intégrer des données cliniques et scientifiques en distinguant bien ce qui est idéal et ce qui est réalisable, tout en tenant compte de l'impact des tests sur le patient
4. Établir un premier plan thérapeutique si jugé nécessaire	<input type="checkbox"/> Le plan ne tient pas compte des données scientifiques et cliniques du patient	<input type="checkbox"/> Le plan suit la logique du diagnostic proposé, mais <input type="checkbox"/> ne tient pas compte du danger potentiel d'erreur	<input type="checkbox"/> Le plan est justifié à la fois sur le plan scientifique et clinique et faisant preuve de prudence réfléchie	<input type="checkbox"/> Le plan est logique avec une argumentation élaborée, tout en s'assurant de la collaboration du patient
Commentaires généraux :	Points à améliorer		Points à maintenir	

Suggestions pédagogiques :

Suivi pédagogique :
Ce qui a été fait depuis un mois :

Ce qui pourrait être fait au cours du prochain mois :

Annexe III Grille d'ÉFPC version initiale

ÉVALUATEUR : _____ DATE : _____

• Résident : _____	Niveau : <input type="checkbox"/> R-I <input type="checkbox"/> R-II <input type="checkbox"/> R-III
• Patient : Type de problème _____	
Niveau de complexité : Faible <input type="checkbox"/> Modéré <input type="checkbox"/> Élevé <input type="checkbox"/>	
• Niveau attendu de performance : R-I = 4 R-II = 5 R-III = 6	

A. Anamnèse

• Le questionnaire médical

1	2	3	4	5	6	7	8	9
Amélioration nécessaire			Questions pertinentes, concises, incluant des données psycho-sociales (valeurs personnelles, niveau socio-économique ...)			Exceptionnel		

Commentaires

• Relation thérapeutique

1	2	3	4	5	6	7	8	9
Amélioration nécessaire			A su établir une relation de confiance, de compréhension, empathique et respectueuse			Exceptionnel		

B. Examen physique

1	2	3	4	5	6	7	8	9
Amélioration nécessaire			Examen pertinent et efficace			A su faire un examen exhaustif de façon efficace, afin de vérifier d'autres diagnostics alternatifs		

C. Synthèse de cas

1	2	3	4	5	6	7	8	9
Amélioration nécessaire			A réussi à bien synthétiser le problème clinique en intégrant les données de l'histoire et de l'examen physique			Exceptionnel		

**D. Diagnostic
proposé**

1	2	3	4	5	6	7	8	9
Doit s'améliorer			A su proposer une solution avec des arguments convainquants en se basant sur des données cliniques recueillies			A su trouver des diagnostics élaborés avec une argumentation claire et précise en se basant sur des données scientifiques intégrant les données cliniques		

Commentaires

**E. Plan
d'investigation**

1	2	3	4	5	6	7	8	9
Doit s'améliorer			Le plan d'investigation est approprié, utile et rentable			A su bien choisir en distinguant ce qui est idéal et ce qui est nécessaire, tenant compte de restrictions médicales et financières		

**F. Plan de traitement
(pour R-III)**

1	2	3	4	5	6	7	8	9
Doit s'améliorer			A su expliquer le traitement proposé au patient			A expliqué de façon claire le traitement et ses effets secondaires		

G. Impression globale et commentaires :

À améliorer :

À maintenir :

H. Durée de l'entrevue médicale :

Durée de la session :

Annexe IV

Grille d'ÉFPC (3 niveaux)

Annexe IV Grille D'ÉFPC trois niveaux

• Évaluateur : _____ Date : _____ Résident : ☐ R-I ☐ R-II ☐ R-III
 • Patient : Type de problème _____ Niveau de complexité : Faible ☐ Modéré ☐ Élevé ☐

Recueil des données cliniques subjectives			Exemples et commentaires	
	À améliorer	Bien	Très bien	
1. Recueil des données cliniques subjectives				
1.1 Effectuer un questionnaire sur le problème médical principal	Les détails cliniques sont obtenus de façon disparate, sans fil conducteur apparent	Les informations obtenues sont concises et pertinentes, faisant preuve d'une connaissance physiopathologique sous-tendant les symptômes rapportés	Des informations principales pertinentes et la chronologie des interventions médicales antérieures (incluant la compliance thérapeutique) ont été obtenues de façon judicieuse en relation aux symptômes rapportés	
1.2 Explorer les aspects psychosociaux de ce même problème médical	La vérification des impacts de la maladie sur la vie du patient est incomplète	L'exploration des habitudes de vie (i.e. : alcool, tabagisme, pratiques sexuelles ...) est faite de façon pertinente à la maladie principale ainsi que l'impact de la maladie	Les impacts de la maladie et les répercussions psychosociales ainsi que les impacts financiers de la maladie sur la vie du patient ont été obtenus judicieusement	
1.3 Utiliser une technique d'entrevue appropriée pour obtenir des informations cliniques	Les questions n'invitent pas des réponses complètes et empêchent l'établissement d'une relation de confiance	A su utiliser différents types de questions (ouvertes, fermées et des techniques tels que résumé, transition ont été utilisés de façon efficace) pour obtenir les informations et clarifier des situations	En plus de l'utilisation judicieuse des types de questions et certaines techniques d'entrevue, a su mettre en confiance le patient par une attitude respectueuse et empathique	
2. Recueil des données cliniques objectives				
	À améliorer	Bien	Très bien	Exemples et commentaires
2.1 Effectuer un examen physique en relation avec le problème médical principal	L'examen physique est fait de façon automatique, non structuré avec des manœuvres parfois inutiles	L'examen physique est pertinent et permet de clarifier des symptômes rapportés. Cet examen a été effectué selon les normes de la spécialité	En plus d'un examen physique efficace, structuré et approprié, a su faire d'autres manœuvres pour vérifier d'autres diagnostics potentiels et alternatifs	
3. Présentation du problème clinique				
	À améliorer	Bien	Très bien	Exemples et commentaires
3.1 Communiquer sa représentation du problème clinique du patient	Le problème clinique présenté ne tient pas compte suffisamment des données cliniques obtenues précédemment	La synthèse du problème clinique a intégré les données subjectives et objectives afin d'arriver à une représentation de problème plausible et compréhensible	L'intégration des données cliniques sont présentées de façon synthétique et claire. De plus, l'ordre des priorités cliniques y est aussi mentionné (selon sa pertinence)	

4. Proposition d'un diagnostic différentiel

	À améliorer	Bien	Très bien	Exemples et commentaires
4.1 Établir un diagnostic différentiel avec une argumentation et une mise en priorité	Le diagnostic proposé est confus, peu logique et dont l'ordre de priorité n'est pas bien explicite.	Les solutions proposées sont logiques et la présentation du problème est fluide, se basant sur des données cliniques	Le diagnostic est logique, clair, élabore avec une argumentation pertinente des données scientifiques tenant compte des données cliniques reçues	

5. Plan d'investigation

	À améliorer	Bien	Très bien	Exemples et commentaires
5.1 Proposer un plan d'investigation pur tenter de solutionner le problème clinique à l'évaluateur	Le plan est inapproprié aux diagnostics proposés et est potentiellement dangereux	Le plan est approprié aux diagnostics proposés et est utile et rentable (a tenu compte des restrictions financières du patient et de l'établissement)	Le plan proposé a su intégrer des données scientifiques et cliniques (ce qui est idéal et ce qui est acceptable par le patient) et a aussi tenu compte des impacts sur le patient	

6. Proposition d'un plan thérapeutique

	À améliorer	Bien	Très bien	Exemples et commentaires
6.1 Établir un plan thérapeutique si jugé nécessaire à la fin de la rencontre	Le plan thérapeutique proposé ne tient pas compte des données scientifiques et cliniques du patient	Les traitements proposés sont justifiés et faisant preuve de la réflexion entre le besoin de traitement immédiat et l'attitude expectative	Les traitements proposés sont argumentés en se basant sur des évidences scientifiques et tenant compte des données cliniques obtenues	

7. Communication d'information au patient

	À améliorer	Bien	Très bien	Exemples et commentaires
7.1 Répondre aux questions du patient	Les termes utilisés sont vagues, jargonneux et difficiles à suivre	Les informations sont données de façon claire et détaillée	Les informations données sont claires, sans jargon avec un souci de s'assurer de la bonne compréhension du patient	

Durée de l'entrevue : début : _____ fin : _____ Durée du feedback : _____
Commentaires généraux : Points à améliorer Points à maintenir

Suggestions pédagogiques :

Prochaine rencontre : ☐ nécessaire ☐ 2 semaines ☐ 4 semaines ☐ 6 semaines ☐ autres

Annexe V

Grille d'ÉFPC (version 4 niveaux)

Annexe V Grille d'ÉFPC (4 niveaux)

Date :

Résident :

Grille remplie par :

Niveau :

☐ R-I

☐ R-II

☐ R-III

I. Recueillir des données cliniques avec le patient

Exemple, commentaires et suggestions

	Inadéquat	Amélioration nécessaire	Adéquat	Excellent	
1. Questionner le patient sur le problème médical principal	<p><input type="checkbox"/> Les détails cliniques sont obtenus de façon disparate, sans fil conducteur apparent</p>	<p><input type="checkbox"/> Les informations sont obtenues de façon concise, et preuve de connaissances physiopathologiques, mais, la chronologie des interventions antérieures n'a pas été obtenue</p>	<p><input type="checkbox"/> Les informations sont obtenues de façon concise, et la chronologie des interventions antérieures est obtenue de façon complète</p>	<p><input type="checkbox"/> Toutes les informations ont été obtenues de façon concise avec un fil conducteur logique faisant preuve de connaissance approfondie de physiopathologie logiques sous-jacentes</p>	
2. Explorer des aspects psycho-sociaux de ce même problème médical	<p><input type="checkbox"/> Aucune vérification des impacts de la maladie sur la vie du patient</p>	<p><input type="checkbox"/> Des habitudes de vie pertinentes ont été obtenues (ex. : tabagisme en relation avec la maladie cardiaque)</p> <p><input type="checkbox"/> Peu a été demandé sur les impacts de la maladie sur la vie du patient</p>	<p><input type="checkbox"/> Les habitudes de vie ainsi que les impacts de la maladie sur la vie du patient ont été adéquatement obtenus</p>	<p><input type="checkbox"/> toutes les informations sur l'impact de la maladie ainsi que la perception du patient vis-à-vis sa maladie ont été obtenues de façon adroite et respectueuse</p>	
3. Utiliser une technique d'entrevue appropriée pour obtenir des informations cliniques nécessaires	<p><input type="checkbox"/> Les questions trop dirigées empêchent des réponses complètes, et ne permettent pas l'établissement d'une relation de confiance</p>	<p><input type="checkbox"/> Les différents types de questions (ouvertes, fermées, clarification, résumé ...) ont été utilisés pour obtenir des informations nécessaires, mais</p> <p><input type="checkbox"/> il a été peu tenu compte de l'état physique du patient (ex. : trop de questions à un patient faible)</p>	<p><input type="checkbox"/> Différents types de questions ont été adéquatement utilisés pour obtenir des informations nécessaires, en tenant compte de l'état clinique du patient, inspirant une relation de confiance</p>	<p><input type="checkbox"/> En plus d'être adroit dans le questionnaire, a toujours été courtois et respectueux et a su établir une relation de confiance qui permet d'obtenir des informations complètes</p>	

Exemple, commentai
et suggestions

	Inadéquat	Amélioration nécessaire	Adéquat	Excellent	
4. Examiner le patient dans différentes circonstances cliniques	<input type="checkbox"/> L'examen physique est fait de façon automatique structurée, avec <input type="checkbox"/> des mobilisations inutiles pour le patient	<input type="checkbox"/> L'examen physique a été fait de façon structurée, mais <input type="checkbox"/> plusieurs manœuvres spécifiques ont été omises faisant preuve d'une compréhension insuffisante de la pathologie du patient	<input type="checkbox"/> L'examen tient compte du problème principal, et <input type="checkbox"/> comporte toutes les manœuvres routinières, et <input type="checkbox"/> raffinées pour identifier le problème du patient, <input type="checkbox"/> tout en respectant le patient, selon les normes de la spécialité	<input type="checkbox"/> L'examen physique est structuré, avec <input type="checkbox"/> un minimum de mobilisation du patient dans une <input type="checkbox"/> durée de temps courte, et <input type="checkbox"/> comporte en plus des manœuvres permettant de vérifier des diagnostics alternatifs. <input type="checkbox"/> De plus, le résident a été respectueux de l'intimité du patient tout au long de l'examen	

5. Discussion d'information avec le patient
5.1 Répondre aux questions des patients avec des informations scientifiques

<input type="checkbox"/> Les termes utilisés sont vagues, jargonneux et difficiles à suivre	<input type="checkbox"/> Les informations données sont claires, même si à quelques reprises, plusieurs jargons médicaux ont été utilisés	<input type="checkbox"/> Les informations sont suffisamment détaillées et claires en tenant compte des données cliniques du patient (a évité des discussions hypothétiques inutiles)	<input type="checkbox"/> Les informations données sont claires, sans jargon avec un réel souci de s'assurer de la bonne compréhension du patient	
---------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------	--

	Inadéquat	Amélioration nécessaire	Adéquat	Excellent	Exemple, commentai et suggestions
--	-----------	-------------------------	---------	-----------	--------------------------------------

1. Présenter le problème clinique du patient à un collègue	<input type="checkbox"/> Le problème présenté n'a pas tenu compte de toutes les informations cliniques, et <input type="checkbox"/> manque de précision dans les détails	<input type="checkbox"/> Le problème présenté tient compte de toutes les informations obtenues, mais <input type="checkbox"/> manque d'intégration trop d'informations inutiles	<input type="checkbox"/> Toutes les informations ont été intégrées pour présenter le problème de façon plausible et compréhensible	<input type="checkbox"/> En plus d'une présentation précise et synthétique des informations cliniques, le résident a établi la priorité clinique pertinente du problème selon une base scientifique (ex. la difficulté respiratoire récente est soulignée avant la perte de poids depuis deux semaines ...)	
------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

	Inadéquat	Amélioration nécessaire	Adéquat	Excellent
2. Proposer un diagnostic différentiel logique	<input type="checkbox"/> Le diagnostic proposé ne tient pas compte des problèmes cliniques rapportés par le patient, et <input type="checkbox"/> dont l'ordre des priorités n'est pas explicite	<input type="checkbox"/> Le diagnostic tient compte du problème clinique, faisant preuve de maîtrise des connaissances de base, mais manque d'argumentation logique	<input type="checkbox"/> Le diagnostic proposé est logique avec une argumentation basée sur des données cliniques de façon juste et précise	<input type="checkbox"/> Le diagnostic est juste et précis, clair avec une argumentation intégrant à la fois les données cliniques et les connaissances scientifiques
3. Proposer un plan d'investigation pour tenter de solutionner le problème clinique du patient	<input type="checkbox"/> Le plan est inapproprié, n'ayant pas tenu compte du diagnostic proposé, et pourrait mettre en danger le patient	<input type="checkbox"/> Le plan est approprié, mais n'a pas tenu compte du patient (ses craintes, ses difficultés à se déplacer...), et comporte des coûts potentiellement excessifs	<input type="checkbox"/> Le plan est approprié en tenant compte de l'impact sur le patient, et des coûts au système de santé	<input type="checkbox"/> Le plan proposé a su intégrer des données cliniques et scientifiques en distinguant bien ce qui est idéal et ce qui est réalisable, tout en tenant compte de l'impact des tests sur le patient
4. Établir un premier plan thérapeutique si jugé nécessaire	<input type="checkbox"/> Le plan ne tient pas compte des données scientifiques et cliniques du patient	<input type="checkbox"/> Le plan suit la logique du diagnostic proposé, mais ne tient pas compte du danger potentiel d'erreur	<input type="checkbox"/> Le plan est justifié à la fois sur le plan scientifique et clinique et faisant preuve de prudence réfléchie	<input type="checkbox"/> Le plan est logique avec une argumentation élaborée, tout en s'assurant de la collaboration du patient

Commentaires généraux : Points à améliorer

Points à maintenir

Suggestions pédagogiques :

Suivi pédagogique :
Ce qui a été fait depuis un mois :

Ce qui pourrait être fait au cours du prochain mois :

